

DISCUSSION PAPER

JULY 2007 ■ RFF DP 07-38

Should Urban Transit Subsidies Be Reduced?

Ian W.H. Parry and Kenneth A. Small

1616 P St. NW
Washington, DC 20036
202-328-5000 www.rff.org



Should Urban Transit Subsidies Be Reduced?

Ian W.H. Parry and Kenneth A. Small

Abstract

This paper derives intuitive and empirically useful formulas for the optimal pricing of passenger transit and for the welfare effects of adjusting current fare subsidies, for peak and off-peak urban rail and bus systems. The formulas are implemented based on a detailed estimation of parameter values for the metropolitan areas of Washington (D.C.), Los Angeles, and London. Our analysis accounts for congestion, pollution, and accident externalities from automobiles and from transit vehicles; scale economies in transit supply; costs of accessing and waiting for transit service as well as service crowding costs; and agency adjustment of transit frequency, vehicle size, and route network to induced changes in demand for passenger miles.

The results support the efficiency case for the large fare subsidies currently applied across mode, period, and city. In almost all cases, fare subsidies of 50 percent or more of operating costs are welfare improving at the margin, and this finding is robust to alternative assumptions and parameters.

Key Words: transit subsidies, scale economies, traffic congestion, welfare effects

JEL Classification Numbers: R48, H21

© 2007 Resources for the Future. All rights reserved. No portion of this paper may be reproduced without permission of the authors.

Discussion papers are research materials circulated by their authors for purposes of information and discussion. They have not necessarily undergone formal peer review.

Contents

1. Introduction.....	1
2. Analytical Model	4
2.1. Model Assumptions	4
2.2. Welfare and Optimal Subsidy Formulas	9
3. Parameter Values	13
4. Results.....	17
4.1. Baseline Results	17
4.2. Sensitivity Analysis	18
4.3. Relation to Recent Studies	20
5. Conclusion	21
References.....	22
Appendix A: Analytical Derivations.....	27
Appendix B. Assessment of Parameter Values.....	35

Should Urban Transit Subsidies Be Reduced?

Ian W.H. Parry and Kenneth A. Small*

1. Introduction

Passenger fares for public transportation are, for the most part, heavily subsidized. Across the 20 largest transit systems in the United States (ranked by passenger miles), the fare subsidy, as measured by the difference between operating costs and passenger fare revenues, ranges from 29 to 89 percent of operating costs for rail, and from 57 to 87 percent for bus (Table 1). Kenworthy and Laube (2001) document a similar pattern of heavy fare subsidies across city transit systems in other developed nations.

Two classic rationales for fare subsidies are often advanced (Glaister 1974; Henderson 1977; Jansson 1979). First, scale economies imply that the marginal social cost of supplying passenger miles is less than the average cost. These scale economies may arise from fixed costs, such as track and station maintenance, but more importantly they arise from the “Mohring effect,” whereby the user’s costs of waiting at transit stops or accessing transit decline as service frequency or route density is increased (Mohring 1972). A related point is that higher passenger density allows vehicles to be operated with higher occupancy, thereby saving on agency costs.

The second rationale is that lower transit fares discourage automobile use, thereby reducing external costs from traffic congestion, local and global air pollution, and traffic accidents. This is a second-best argument, since it assumes that these external costs cannot be internalized through appropriate road pricing.

Determining whether current fare subsidies are warranted by these two arguments is complicated by several factors. First, the strength of both arguments may vary greatly by time of day, mode, and location. Second, the appropriate subsidy depends on how transit agencies respond to increases in passenger demand at the margin—whether by expanding service through more vehicle miles (thus providing higher service frequency and/or a denser route network) or by increasing vehicle occupancy (either through higher load factors and/or larger vehicles). Third,

* We are grateful to Richard Arnott, Bruno De Borger, Stef Proost, George Tolley, and Kurt Van Dender for helpful comments and suggestions and to Elliot Klein for research assistance. Kenneth Small thanks the University of California Energy Institute for financial support.

transit vehicles themselves may contribute to externalities such as congestion and pollution, and their passengers generate external costs on each other via crowding (Kraus 1991) or increased boarding and alighting time. Fourth, automobile externalities are partly internalized through fuel taxes. And finally, altering the subsidy for one mode will cause substitution effects across modes and times of day, with secondary effects on economic efficiency due to the many distortions from optimality conditions in the system.

Several studies have estimated optimal transit prices, focusing on one or both of the primary rationales just mentioned and usually just one location. None of them encompass all of the additional complications identified above. In fact, existing estimates of optimal transit prices (given current road prices) vary enormously, from zero to more than 100 percent of operating costs, providing a confusing guide as to whether current fare subsidies should be preserved, expanded, or eliminated.¹ It is difficult to discern the reasons for the strikingly diverse results because the studies apply to different regions and years, they account for different factors, they make different assumptions about transit agency response, not all of them distinguish time of day, and some use simplified analytical models while others use less transparent but more sophisticated computational models. Furthermore, assumptions about agency and travel responses that may be reasonable at current prices may not be at very different prices; thus, for the purpose of drawing robust qualitative results, it may be better to focus on the direction of welfare effects from small changes to existing prices rather than place too much emphasis on fully optimized prices.

This paper provides a general framework for evaluating existing fare subsidies and potential pricing reforms. It does so by developing a single analytical model that incorporates all the factors just described, then derives formulas for optimal subsidies for bus and rail at peak and off-peak periods and calculates the welfare effects from incrementally adjusting current fare subsidies. These formulas clarify the contribution of all underlying parameters and can be

¹ For London, Glaister and Lewis (1978, Table 4, line 3b) estimate optimal rail and bus fares at about 50 to 60 percent of marginal operating costs. For the San Francisco Bay Area and for Pittsburgh, Viton (1983) finds optimal fares to be virtually zero. Winston and Shirley (1998) find quite the opposite for the United States as a whole, with optimal bus and rail fares covering 84 percent and 97 percent of marginal operating costs, respectively. For a prototype Belgian city, De Borger et al. (1996) estimate optimal transit fares are 50 to 114 percent of average agency costs, depending on how service frequency adjusts to passenger demand. For Brussels, Van Dender and Proost (2004) estimate optimal transit fares to be nearly zero in peak periods and about double current fares in off-peak periods. Two recent studies of Washington, D.C., by Winston and Maheshri (2007) and Nelson et al. (2007), estimate net total benefits from transit but with conflicting results. Insofar as possible, we relate our findings to this previous literature in Section 4.3.

empirically implemented in a spreadsheet. Following an extensive compilation and estimation of parameter values, we apply the formulas to three large but very different metropolitan areas: Washington, Los Angeles, and London. Our analysis includes vehicle capital costs (which can be varied fairly easily) but not infrastructure investments; thus, following previous optimal transit pricing literature, we explore how best to use existing infrastructure without worrying about recovering sunk capital costs.

The most striking finding is that, in almost all cases, extending fare subsidies beyond 50 percent of operating costs—often well beyond—is welfare improving at the margin across modes, periods, and cities. And these findings are generally robust to plausible alternative assumptions about parameters and agency behavior. The main reasons why large subsidies are welfare improving are the two classic ones. However, the relative importance of these two rationales varies across different cases and assumptions. We find big gains from diverting auto traffic, especially during peak periods. Furthermore, to the extent that service is increased in response to additional passenger demand, scale economies arising from reduced user costs of wait and access are usually significant, especially for bus and for off-peak service. And to the extent that transit vehicle occupancy is increased instead, savings in operating costs typically outweigh any extra costs from crowding or increasing vehicle size.

Do our results imply that existing operating deficits should continue to be financed through general taxation, rather than being reduced through substantial increases in passenger fares and reduced service levels?

One counterargument is that we ignore the broader efficiency costs from financing operating deficits through distortionary taxes. However, as emphasized in the literature on environmental tax shifts (e.g., Bovenberg and Goulder 2002; Parry and Bento 2001), there are important counteracting effects on tax distortions elsewhere in the economy to the extent that lower transportation costs encourage more economic activity. We discuss tax distortions in Section 5; based on a rough calculation there, the net impact of these distortions on optimal subsidies appears to be moderate.

Another issue is to what extent our results may carry over to urban transit systems other than the three studied here. Marginal congestion costs are likely to be lower in most other cities, but as discussed in our sensitivity analysis, optimal fare subsidies can still be substantial because of other factors. A more definitive answer awaits a detailed parameter assessment for other cases.

Probably the most important qualification is that we do not explicitly model the potentially lax incentives for cost minimization inherent in a publicly provided service. As shown later, our general framework and results still apply to more efficiently managed transit systems with lower operating costs. However, there is evidence that subsidy programs themselves cause cost-inflation through excessive compensation, misuse of high-skilled labor in low-skill tasks, and inefficient use of labor and capital (Winston and Shirley 1998; Small and Gomez-Ibanez 1999, sect. 3.3). One response to this problem might be to privatize transit systems while retaining some subsidies; but an alternative would be to switch to a fixed subsidy per passenger mile (by mode and period) and require the agency to cover the remainder of its operating costs through its pricing structure.

The rest of the paper is organized as follows. Section 2 describes the analytical model and derives key formulas; Section 3 discusses baseline data; Section 4 presents the main quantitative results and sensitivity analysis; and Section 5 concludes and elaborates on the qualifications.

2. Analytical Model

We develop a model of urban passenger travel by auto, rail, and bus at different times of day, where transit user costs depend on congestion, transit frequency, route density, and vehicle crowding. Travel also produces pollution and accident externalities, some of which are internalized by fuel taxes. The government chooses transit characteristics and fares subject to a budget constraint, while agents optimize over travel choices taking externalities and transit characteristics as given.

2.1. Model Assumptions

(i) *User utility.* The representative agent has preferences defined as follows:

$$(1a) \quad U = u(X, M, \Gamma) - Z$$

$$(1b) \quad M = M(\{M^{ij}, i = P, O; j = CAR, B, R\})$$

$$(1c) \quad \Gamma = \Gamma(T, W, A, C)$$

$$(1d) \quad T = \sum_{ij} t^{ij} M^{ij}, \quad W = \sum_{ij \neq iCAR} w^{ij} M^{ij}, \quad A = \sum_{ij \neq iCAR} a^{ij} M^{ij}, \quad C = \sum_{ij \neq iCAR} c^{ij} M^{ij}$$

where all variables are in per capita terms. In (1a), X is the quantity of a numeraire or general consumption good; M is subutility from passenger miles traveled; Γ is a generalized (non-

money) cost of travel; and Z is disutility from pollution and traffic accident externalities.² In (1b), M^{ij} is passenger miles traveled during period i by mode j where the two time periods are $i = P$ (peak) and O (off-peak), and the three modes are $j = CAR$ (auto), B (bus), and R (rail).³ In (1c), T is total in-vehicle travel time, W is time spent waiting at transit stops, A is time spent accessing transit, and C is crowding experienced on transit; as shown in (1d), these non-money costs are an aggregation over miles traveled, each multiplied by the respective per mile costs t^{ij} , w^{ij} , a^{ij} , and c^{ij} . We assume $u(\cdot)$ is increasing and quasi-concave in X and M and decreasing and quasi-concave in Γ ; $M(\cdot)$ is quasi-concave, implying travel by different modes and time of day are imperfect substitutes; and $\Gamma(\cdot)$ is increasing and quasi-concave in non-monetary travel inputs.

(ii) *Travel characteristics.* Several characteristics of transit vehicles affect user and operator costs. First is vehicle *occupancy*, o^{ij} , the average number of passengers in a bus or train:

$$(2a) \quad o^{ij} = M^{ij} / V^{ij}$$

where V^{ij} is total vehicle miles. Second is the *load factor*, l^{ij} , defined as the fraction of a vehicle's passenger capacity n^{ij} that is occupied:

$$(2b) \quad l^{ij} = o^{ij} / n^{ij}$$

Third is the average *service frequency*, f^{ij} , along each bus or rail transit line:

$$(2c) \quad f^{ij} = V^{ij} / D^{ij}$$

where D^{ij} is route density, measured as total route miles within the fixed service area.⁴

These variables determine the per-mile travel characteristics in (1d) as follows:

$$(3a) \quad t^{ij} = t^j (V^{iCAR} + \alpha_B V^{iB}) + \theta^j o^{ij}, \quad j = CAR, B; \quad \theta^{CAR} = 0; \quad t^{iR} = t^R + \theta^R o^{iR}$$

$$(3b) \quad w^{ij} = w^{ij}(f^{ij}), \quad a^{ij} = a^{ij}(D^{ij}), \quad c^{ij} = c^{ij}(l^{ij}), \quad j=B,R$$

where $\alpha_B > 1$ is the contribution of a bus to congestion relative to that of a car, or the “passenger

² We exclude possible externalities from oil dependence because they are difficult to define; insofar as they have been quantified (for example, NRC 2002 put them at 12 cents per gallon of gasoline), incorporating them would make very little difference to our results, as can be seen from our discussion of sensitivity with respect to global warming damages in Section 5. Also, some of the costs of oil dependence are domestic rather than worldwide, and so including all of them would be inconsistent with the worldwide perspective adopted in estimating global warming costs.

³ We hold trip length constant, so variations in M^{ij} arise from variations in the number of trips.

⁴ As is normal in economic models, all variables are flows. Thus, for example, V^{PR} is defined as rail vehicles *per hour* averaged over the peak period. Hence V^{PR}/D^{PR} has units (vehicle/hour)/route; that is, vehicles per hour along a given route.

car equivalent.” In (3a), in-vehicle time has two potential components. First is the time transit vehicles are stationary at transit stops, expressed per passenger mile; this is equal to vehicle occupancy times θ^j , which is the average dwell time per passenger (boarding plus alighting) divided by trip length. Second is the time the vehicle spends in motion per mile of travel, $t^j(\cdot)$, which is the inverse of vehicle speed. For autos and buses, which share the roads, $t^j(\cdot)$ is a weakly convex function of aggregate road traffic, with bus traffic weighted by α_B ; buses travel more slowly than autos, therefore $t^B(\cdot) > t^{CAR}(\cdot)$. For rail we assume t^R is fixed; that is, an extra train does not slow down the speed of other trains in the system.

In (3b) the per mile wait time for transit varies negatively with service frequency; the per mile transit access time varies negatively with route density; and per mile crowding varies positively with the load factor.

(iii) *Pollution and accident externalities.* The nature of these externalities has been discussed extensively elsewhere (e.g., Parry and Small 2005); we simply summarize their aggregate cost by

$$(4) \quad Z = \sum_{ij} z^{ij} V^{ij}$$

where z^{ij} is the combined pollution and accident external costs per vehicle mile.⁵

(iv) *Household optimization.* The household budget constraint is

$$(5) \quad I - TAX = X + \sum_{ij} p^{ij} M^{ij}$$

where I is (exogenous) private income, TAX is a lump-sum tax to help finance transit deficits, p^{ij} is the money cost per passenger mile of travel, and the price of X is normalized to one. For bus and rail, p^{ij} is the per mile fare, whereas for auto, $p^{iCAR} = \bar{p}^{iCAR} + \tau^{iCAR}$, where \bar{p}^{iCAR} is pretax fuel costs and τ^{iCAR} is fuel taxes, both expressed per passenger mile (\bar{p}^{iCAR} and τ^{iCAR} vary by time of day because congestion affects fuel economy).⁶

Households choose passenger miles and the numeraire good to maximize utility (1)

⁵ Some of the social costs of traffic accidents (e.g., injury risk to oneself) are internal and are implicitly taken into account in the subutility function $M(\cdot)$ for travel.

⁶ Other money payments (e.g., car maintenance, parking fees) are assumed constant and are treated as subtractions from the utility of traveling by car rather than explicitly as costs.

subject to (5), taking p^{ij} , t^{ij} , w^{ij} , a^{ij} , c^{ij} , Z , and TAX as given. This yields first-order conditions, summarized by

$$(6a) \quad \frac{u_{M^{ij}}}{u_X} = q^{ij} \equiv p^{ij} + \rho^T t^{ij} + \rho^W w^{ij} + \rho^A a^{ij} + \rho^C c^{ij}$$

$$(6b) \quad \rho^k \equiv -u_{\Gamma} \Gamma_k / u_X, \quad k=T, W, A, C$$

The quantities ρ^k are the (marginal) dollar values of in-vehicle time, waiting time, access time, and crowding, which are taken as fixed (although it is not indicated by the notation, we allow these values to vary by time of day.) Thus q^{ij} is a generalized price, including both money and nonmoney costs per mile; agents equate the marginal benefit from passenger miles $u_{M^{ij}} / u_X$ to this generalized price for each mode and time period. From (5), (6), and (1) we obtain the demand functions and indirect utility (denoted by \sim):

$$(7) \quad M^{ij} = M^{ij}(\{q^{xy}\}, TAX), \quad X = X(\{q^{xy}\}, TAX), \quad \tilde{U} = \tilde{u}(\{q^{xy}\}, TAX) - Z$$

where $\{q^{xy}\}$ denotes the set of q^{ij} for i, j .

(v) *Transit agency constraints.* The agency's total operating cost, OC^{ij} , in period i for mode j , is

$$(8a) \quad OC^{ij} = F^{ij} + K^{ij} t^{ij} V^{ij}$$

$$(8b) \quad K^{ij} = k_1^{ij} + k_2^{ij} n^{ij}$$

where k_1^{ij} , $k_2^{ij} > 0$ are parameters. In (8a), F^{ij} is a fixed cost representing, for example, the cost of operating rail stations; consistent with empirical evidence, we assume there are no scale economies or diseconomies in providing bus vehicle miles, $F^{iB}=0$ (Small 1992, 57). Operating costs also include variable costs equal to total vehicle hours of operation $t^{ij} V^{ij}$ multiplied by variable costs per vehicle hour, K^{ij} , which primarily reflect driver labor and vehicle capital. In (8b), K^{ij} is a linear function of vehicle capacity, with scale economies to the extent $k_1^{ij} > 0$. We assume $k_1^{Pj} > k_1^{Oj}$ because peak service does not conveniently fit an eight-hour workday, so its unit labor costs are higher; and we assume $k_2^{Pj} > k_2^{Oj}$ because larger vehicles that are purchased primarily for peak use are also available off-peak at little or no extra cost.

The agency budget constraint is

$$(9) \quad TAX + \sum_i \tau^{iCAR} V^{iCAR} = \sum_i \sum_{j \neq iCAR} (OC^{ij} - p^{ij} M^{ij})$$

That is, revenues from lump-sum taxes and fuel taxes finance the transit deficit.⁷

(vi) *Agency adjustment of transit characteristics.* Because there is little empirical basis for quantifying access and crowding costs, we eliminate the need to do so by assuming that, for given vehicle miles, the transit agency optimizes over route density and service frequency, and that for a given vehicle occupancy, it optimizes over vehicle size and load factor. These assumptions imply the following first-order conditions (see Appendix A):

$$(10a) \quad \rho^W w^{ij} \eta_w^{ij} = \rho^A a^{ij} \eta_a^{ij}$$

$$(10b) \quad \rho^C c^{ij} \eta_c^{ij} o^{ij} = t^{ij} k_2^{ij} n^{ij}$$

where η_w^{ij} , η_a^{ij} , η_c^{ij} denote wait, access, and crowding cost elasticities, all defined positively: for example, $\eta_w^{ij} = |dw^{ij}/df^{ij}| \cdot (f^{ij}/w^{ij})$. (10a) states that route density is increased until the incremental cost of extra waiting, resulting from less frequent service, equals the incremental reduction in access cost. (10b) states that transit vehicle size is increased until the incremental reduction in crowding costs to its occupants equals the incremental agency cost of the larger vehicle. Although these assumptions represent a neutral case,⁸ we discuss later the implications of relaxing them. From (6a) and (10) we can express the generalized user price as

$$(10c) \quad q^{ij} = p^{ij} + \rho^T t^{ij} + \rho^W w^{ij} \cdot (1 + \eta_w^{ij} / \eta_a^{ij}) + t^{ij} k_2^{ij} n^{ij} / (o^{ij} \eta_c^{ij})$$

⁷ In practice, fuel tax revenues are earmarked for road and transit infrastructure projects; accounting for this could affect our results very slightly, to the extent that the social benefit per dollar of infrastructure spending differs from unity.

⁸ For example, if the agency overinvests in service frequency relative to route density, then using (10a) and data on wait costs will underestimate (unobserved) marginal access costs, and vice versa if there is underinvestment in service frequency. Without reliable data on access costs, we cannot say which of these two cases might be the more likely.

Condition (10a) ignores fixed costs of additional routes and so overstates the optimal route density for rail. On the other hand, there is evidence that rail lines have been built that are not economically justified, and thus current route density may also exceed optimal route density. Furthermore, off-peak route density can be adjusted even in the short run by making some lines peak-only. To analyze this more thoroughly, we could omit (10a) for peak service and assume instead that D^{Pj} is fixed at its current value; this would require empirical estimates of crowding costs $\rho^C c^{Pj}$.

Following an increase in demand for passenger miles, we assume that a (constant) fraction ε_V of it is accommodated through increased vehicle miles (of which the increases in service frequency and route density are chosen to satisfy (10a)) and fraction $1-\varepsilon_V$ through higher occupancy of transit vehicles (with the increase in vehicle size and load factors chosen to satisfy (10b)).

2.2. Welfare and Optimal Subsidy Formulas

(i) *Marginal welfare effects.* We first consider welfare effects of marginal changes in existing transit prices. The resulting formulas are useful in drawing robust conclusions about whether increasing current fare subsidies improves or reduces economic efficiency—robust because the formulas depend only on marginal rather than global assumptions about demand functions and agency adjustments. We focus on peak-period rail for exposition; the formulas for other transit modes and periods are analogous.

We differentiate indirect utility with respect to $-p^{PR}$; that is, we consider an incremental reduction in the fare, accounting for induced changes in travel, user, and external costs, and in the agency budget. The resulting marginal welfare effect, defined (in consumption units) as $MW^{PR} \equiv -(d\tilde{U} / dp^{PR}) / u_x$, can be expressed as the sum of four components (see Appendix A):

$$(11) \quad MW^{PR} \equiv \underbrace{-(MC_{supply}^{PR} - p^{PR})(-M_{PR}^{PR})}_{\text{marginal cost/price gap}} + \underbrace{(MB_{scale}^{PR} - MC_{occ}^{PR})(-M_{PR}^{PR})}_{\text{net scale economy}} + \underbrace{\sum_{ij=PR, iCAR} MC_{ext}^{ij} \cdot M_{PR}^{ij}}_{\text{externality}} \\ + \underbrace{\sum_{ij=OR, PB, OB} (MC_{supply}^{ij} + MC_{ext}^{ij} + MC_{occ}^{ij} - MB_{scale}^{ij} - p^{ij}) M_{PR}^{ij}}_{\text{other transit}}$$

In (11), the quantity $M_{PR}^{ij} \equiv dM^{ij} / dp^{PR}$ is the marginal demand shift for mode ij induced by a peak-rail price change. Our preference assumptions imply that $M_{PR}^{PR} < 0$ and $M_{PR}^{ij} \geq 0$ for $ij \neq PR$; that is, peak-rail ridership goes up following a decrease in the fare, diverting ridership away from autos and other transit.

The other expressions in (11) are defined as follows:

$$(12a) \quad MC_{supply}^{ij} = (\varepsilon_V / o^{ij}) K^{ij} t^{ij}$$

$$(12b) \quad MB_{scale}^{ij} = \varepsilon_V \rho^W w^{ij} \eta_w^{ij}, \quad MC_{occ}^{ij} = (1 - \varepsilon_V) t^{ij} k_2^{ij} n^{ij} / o^{ij}$$

$$(12c) \quad MC_{ext}^{iCAR} = \left(\frac{z^{iCAR}}{u_X} + MC_{cong}^{iCAR} - \tau^{iCAR} \right) / o^{iCAR}$$

$$MC_{ext}^{ij} = \varepsilon_V \left(\frac{z^{ij}}{u_X} + MC_{cong}^{ij} \right) / o^{ij} + (1 - \varepsilon_V) MC_{dwell}^{ij}, \quad j=B,R$$

$$(12d) \quad MC_{cong}^{iCAR} = \sum_{k=CAR,B} t_{CAR}^{ik} \rho^T M^{ik} + t_{CAR}^{iB} K^{iB} V^{iB}; \quad MC_{cong}^{iB} = \alpha_B MC_{cong}^{iCAR}; \quad MC_{cong}^{iR} = 0$$

$$MC_{dwell}^{ij} = \theta^j \cdot (\rho^T o^{ij} + K^{ij})$$

In (12a), MC_{supply}^{ij} is the marginal cost to the transit agency of supplying an extra passenger mile; it equals the product of the travel time per mile, the variable operating cost per unit of time, and the response of vehicle miles to an extra passenger mile, ε_V / o^{ij} . Compared with (8a), the marginal supply cost is likely to be *below* the average operating cost per mile, to the extent that $\varepsilon_V < 1$ and/or there are fixed costs.

In (12b), MB_{scale}^{ij} is the marginal user benefit per extra passenger mile from scale economies. It is positive to the extent that vehicle miles respond to passenger miles, $\varepsilon_V > 0$; it includes the reduction in wait costs from increased service frequency and the reduction in access costs from the increase in route density, with the latter included as a wait cost equivalent using (10a). MC_{occ}^{ij} is the marginal cost of increased vehicle occupancy per extra passenger mile and is positive if $1 - \varepsilon_V > 0$. It incorporates the increase in agency supply costs from increased vehicle size and the increase in crowding costs from higher load factors, with the latter expressed as an agency cost equivalent using (10b).

In (12c), MC_{ext}^{ij} denotes net external costs per passenger mile. For autos, it equals the per-vehicle-mile external cost of pollution, accidents, and congestion (the latter denoted MC_{cong}^{ij}), net of the fuel tax, and all divided by occupancy to convert to passenger miles. For transit, MC_{ext}^{ij} includes these same costs to the extent that vehicle miles respond to passenger miles ($\varepsilon_V > 0$), except there are no congestion costs for rail; in addition, it includes the marginal cost of increased dwell time, MC_{dwell}^{ij} , applicable to the extent that vehicle occupancy increases ($1 - \varepsilon_V > 0$). Fuel taxes for transit are excluded from supply costs and thus do not need to be netted out here.

In (12d), each of MC_{cong}^{iB} and MC_{cong}^{iCAR} measures the increase in travel time to all highway users from an extra passenger mile by bus or auto, scaled by the value of travel time, plus the increase in bus operating costs because it takes longer (and therefore requires more labor and capital input) to supply a passenger mile with slower-moving traffic. Finally, MC_{dwell}^{ij} is the effect on other passengers' time costs, and on agency operating costs, due to the additional boarding and alighting time when an extra passenger mile is accommodated through higher occupancy.

Revisiting (11), each term shows a component of welfare change due to shifting from other modes and/or time periods into peak rail. The “marginal cost/price gap” term shows that welfare from a price reduction is reduced to the extent that the fare for peak rail already falls short of the corresponding marginal supply cost. The “net scale” term indicates that welfare from a fare reduction is larger to the extent that scale economies from increased peak-rail use outweigh the extra user costs due to crowding and the extra user and agency costs from increased occupancy of peak-rail vehicles. The “externality” term shows that welfare increases insofar as pollution, accident, and congestion externalities from auto travel are reduced, although this is partly offset if there are similar externalities from peak rail itself. Finally, the “other transit” term indicates that welfare improves to the extent that passengers are diverted from other transit modes or times of day whose fares fall short of the corresponding marginal social cost; that marginal social cost includes incremental supply cost, occupancy cost, and externalities, less incremental benefits from scale economies.

(ii) *Optimized transit subsidies.* Equation (11) gives us all we need to compute marginal welfare change from increasing an existing subsidy. If we want to go further and find the optimal subsidy, we can do so by setting (11) to zero, with the qualification that we have less confidence in measuring its components when prices are far from current values. Doing so, we obtain the following result for optimal fare subsidy per passenger mile, \hat{s}^{PR} :

$$(13) \quad \hat{s}^{PR} = OC^{PR} / M^{PR} - \hat{p}^{PR} = \underbrace{OC^{PR} / M^{PR} - MC_{supply}^{PR}}_{\text{average/marginal cost gap}} + \underbrace{(MB_{scale}^{PR} - MC_{occ}^{PR})}_{\text{net scale economy}} \\ + \underbrace{\sum_i MC_{ext}^{iCAR} \cdot m_{PR}^{iCAR} - MC_{ext}^{PR}}_{\text{externality}} + \underbrace{\sum_{ij=OR, PB, OB} (MC_{supply}^{ij} + MC_{ext}^{ij} + MC_{occ}^{ij} - MB_{scale}^{ij} - p^{ij}) m_{PR}^{ij}}_{\text{other transit}}$$

where \hat{p}^{PR} is value of p^{PR} that sets (11) to zero and $m_{PR}^{ij} = -M_{PR}^{ij} / M_{PR}^{PR}$ is the modal diversion ratio, or fraction of increased travel by peak rail that comes from reduced travel by model j in period i . Equation (13) implies that the optimal subsidy per passenger mile is positive to the extent that (a) marginal supply cost is below average operating cost; (b) scale economies from increasing passenger miles outweigh costs from increased occupancy; (c) externality gains from diverting auto travel exceed the marginal external costs of the increased peak-rail travel; and (d) travel is diverted from other transit for which the overall social cost per passenger mile exceeds the fare.

As already discussed, equation (13) may be reliable only when conditions are not too far from those currently observed. However, we found that attempts to simultaneously optimize all transit fares sometimes led to drastic changes in ridership and consequently in transit characteristics. Therefore in the empirical simulations presented here, we optimize over a single transit price while setting prices of competing transit modes and periods at their currently observed levels. In other words, we ask what a given fare should be, given the possibly nonoptimal levels of other fares.

(iii) *Functional forms.* We assume that marginal congestion costs MC_{cong}^{iB} and MC_{cong}^{iCAR} are constant because road traffic changes only moderately in our policy simulations; we also assume that z^{ij} / u_x and η_c^{ij} are constant but that η_w^{ij} and η_a^{ij} vary as discussed in Section 3.

Passenger travel demands are assumed to have constant elasticities with respect to own generalized price, and to adjust to other prices according to the modal diversion elasticities already defined. Writing this out for changes in the price of peak rail, we obtain

$$(14a) \quad M^{PR} = \bar{M}^{PR} \left(\frac{q^{PR}}{\bar{q}^{PR}} \right)^{\eta_q^{PR}},$$

$$(14b) \quad M^{ij} = \bar{M}^{ij} - \int_{\bar{p}^{PR}}^{p^{PR}} m_{PR}^{ij} M_{PR}^{PR} dp^{PR}, \quad ij \neq PR$$

where a bar denotes an initial (currently observed) value, and η_q^{PR} is the elasticity of demand for peak rail with respect to its generalized price. Differentiating (14a), we obtain explicitly the dependence of peak-rail demand on its money price, holding the generalized prices of other

modes constant:

$$(14c) \quad M_{PR}^{PR} \equiv \frac{dM^{PR}}{dp^{PR}} = \eta_q^{PR} \frac{dq^{PR}}{dp^{PR}} \frac{M^{PR}}{q^{PR}}$$

In (14c), dq^{PR} / dp^{PR} is the total effect of a one-cent-per-mile increase in the passenger fare on the generalized cost of peak-rail travel, through equation (10c); that effect is greater than one cent because the reduction in peak-rail vehicle miles increases wait and access costs per mile (assuming $\varepsilon_V > 0$), which magnifies the depressing effect on ridership.⁹

3. Parameter Values

We focus on areas served by the Washington Metropolitan Area Transit Authority (WMATA), the Los Angeles County Metropolitan Transit Authority (MTA), and Transport for London (TfL) for year 2002. Appendix B provides an extensive discussion of data sources and various estimation procedures for all parameters. Below we comment on selected baseline data summarized in Table 2; alternative assumptions with possible significance for our results are discussed later.

(i) *System aggregates and agency adjustment.* The Washington and Los Angeles transit systems each carry nearly 2 billion passenger miles a year across all modes and times of day; this transit usage represents 4.3 percent of total passenger miles (auto plus transit) in Washington but only 1.3 percent in Los Angeles. In London, the transit system carries more than 8 billion passenger miles a year, or 21.7 percent of all passenger travel. For Washington, passenger miles by rail are more than three times those for bus, while the opposite applies to Los Angeles, with its extensive bus but limited rail network. For London, the two modes are closer in size, with passenger miles for rail exceeding those for bus by 29 percent. Average transit vehicle occupancies are broadly comparable across the cities but are 26 to 76 percent greater during peak than during off-peak periods. Train occupancy is around 5 to 10 times that for bus.

⁹ We make this point especially because many of the studies empirically measuring money-price elasticities of transit demand have not held wait and access costs constant while observing changes in money price. Thus the elasticity they measure involves the total money-price derivative defined by (14c) rather than a partial derivative that holds service characteristics constant.

We assume that transit agencies meet a 1 percent increase in passenger demand through a 0.67 percent increase in vehicle miles and a 0.33 percent increase in vehicle occupancy, or $\varepsilon_V = 0.67$. As explained in Appendix A, this rule would apply, under certain simplifications, if the agency optimally trades off vehicle miles and occupancy and if wait and access times are inversely proportional to frequency and route density, respectively.¹⁰ We consider other values for ε_V in our sensitivity analysis.

(ii) *Operating costs, marginal supply costs, and fares.* Our cost data enable us to compute the parameters in (8). They imply that average operating costs per *vehicle* mile are around 60 to 100 percent larger in the peak than in the off-peak period. Peak costs are greater because they include vehicle capital costs, higher unit labor costs due to irregular work hours, and in the case of bus, additional costs incurred because it takes longer to drive a mile on congested roads.

However, the peak/off-peak discrepancies in the average operating costs per *passenger* mile are much smaller (approximately zero for rail) because of the different vehicle occupancies. The resulting figures for average operating costs vary from 30 to 103 cents per passenger mile across modes, periods, and time of day. For the U.S. cities, average operating costs per passenger mile are generally higher for bus than for rail, particularly for Washington, where bus occupancies are lower than in Los Angeles. The opposite applies for London. There, the *marginal* cost of supplying passenger miles, from equation (12a), is two-thirds of the average costs in the case of bus and only 60 percent of average costs in the case of rail, because $\varepsilon_V = 0.67$ and 10 percent of average rail costs is assumed fixed.

Passenger fares are 20 to 25 cents per mile for Washington and London; in Los Angeles they are only 14 cents per mile for bus and 8 cents per mile for rail.¹¹ Fare subsidies, defined as $(OC^{ij} - p^{ij}M^{ij})$, are substantial and exceed 50 percent or more of average operating costs in almost all cases; subsidies are especially large for Los Angeles rail (82 to 83 percent), with its unusually low fares, and also for Washington bus (73 to 81 percent), which has typical fares but relatively low occupancies.

¹⁰ See also Nash (1988), Jansson (1997), and Small (2004). The result is a modification of the better-known “square-root rule” (Mohring 1972) that applies when route density, and hence access costs, are fixed.

¹¹ The low rail fare in Los Angeles was so pronounced that it resulted in a suit by a bus riders’ group against the operating agency in 1996; however, this resulted in lowering the bus fare rather than raising the rail fares to levels comparable to those in other cities.

(iii) *User costs.* Average wait times at transit stops are estimated from service frequency. We assume that when vehicles are less than 15 minutes apart, travelers arrive at random, so the wait-time elasticity is one; but that as the time between vehicles exceeds 15 minutes, an increasing fraction of travelers use a timetable, thereby lowering the elasticity (see Appendix B). Expressing wait times on a per mile basis and multiplying by the value of wait time ρ^W (assumed from the empirical literature to be 60 to 80 percent of the market wage, depending on time period) give initial wait costs that vary from 5 to 72 cents per passenger mile. Wait times are much larger during the off-peak period; they are also larger for bus than for rail.

There is less empirical basis for gauging crowding and access time elasticities; we have assumed location-specific values as explained in Appendix B. At least when equation (10) applies, our results are not very sensitive to alternative assumptions about these elasticities.

(iv) *Marginal benefit from scale economies and marginal occupancy costs.* These are computed from (12b) using our parameters for wait costs and vehicle capital costs.

Because of greater wait times at transit stops, marginal scale economies are larger for bus than for rail and for off-peak than for peak travel; however, for bus travel, wait times are less responsive to service frequency in off-peak periods when some people are using schedules, which narrows the discrepancy in scale economies across time of day for that mode. Overall, marginal scale economies vary between 3 and 34 cents per mile across modes, periods, and cities. Increased occupancy costs per additional passenger mile counteract some, though usually not all, scale economies at peak period; however, they are zero in the off-peak period because all vehicle capital costs (and hence crowding costs) are attributed to the peak.

(v) *Externalities.* Marginal external costs per passenger mile for autos are dominated by congestion; that is, the net impact of pollution and accident externalities and fuel taxes is relatively modest. This is particularly the case for London, where marginal congestion costs are 103 and 37 cents per passenger mile in the peak and off-peak periods, respectively, but global and local pollution and accident costs are “only” about 4 cents per passenger mile, with offsetting fuel taxes of 6 to 9 cents per mile. Perhaps the most contentious assumption is that global warming costs amount to less than half of one cent per mile, though that is what most mainstream estimates imply; even increasing our estimate several-fold would still leave pollution

costs small relative to congestion costs.¹² For the U.S. cities, overall external costs for auto are 25 to 31 cents and 6 to 9 cents per passenger mile, respectively, in the peak and off-peak periods; figures for gridlocked London are much higher, at 99 and 35 cents, respectively.¹³

Accident and pollution costs for bus are minimal per passenger mile because of the sufficiently high vehicle occupancies; the marginal costs of increased dwell time are also not very large. However, marginal congestion costs are more substantial and amount to 15 to 29 cents per passenger mile for London (assumed passenger car equivalents for bus are between 4 and 5). Marginal external costs for rail are negligible, since we assume no congestion.

(vi) *Travel responses.* Based on literature surveys of transit demand elasticities (see Appendix B), we choose peak and off-peak own-fare elasticities of -0.24 and -0.48 for rail and -0.4 and -0.8 for bus. Elasticities with respect to generalized prices, η_q^{ij} , are then obtained using (10c), assuming that the empirical estimates of own-fare elasticities incorporate the indirect effects dq^{ij}/dp^{ij} as embodied within (14c).

Modal diversion ratios are based on available evidence and our own judgment (Appendix B). We assume that 60 to 85 percent of increased passenger mileage in response to lower fares comes from diverted auto travel for U.S. cities, and 40 to 50 percent for London, where autos account for a smaller share of passenger travel (Table 2). We assume that 10 percent of extra travel on one transit mode comes from the same mode in the other period, and that the fraction from the other transit mode within the same time period is 5 percent for Los Angeles, 10 percent for Washington, and 30 percent for London.

¹² A gallon of gasoline contains 0.0024 tons of carbon, so even an extremely large carbon price of, say, \$200 per ton amounts to 48 cents per gallon, or about 3 cents per auto mile for peak periods in the United States and even less for other cases.

¹³ Our figures are measured prior to the introduction of the London congestion toll in 2003. However, given its very limited geographical coverage, we would expect it to have only a modest effect on marginal congestion cost across the entire city. If our study were more disaggregated geographically, congestion charging might have more significant effects on optimal transit fares within central London, both by reducing congestion and by increasing the value of τ^{iCAR} there.

4. Results

4.1. Baseline Results

The upper part of Table 3 shows estimates of the marginal welfare effect of a one-cent-per-mile reduction in the passenger fare, starting either at the current subsidy level or at a subsidy level equal to 50 percent of operating costs. Results are expressed in U.S. cents per passenger mile (at 2004 price levels).

The most striking result is that, with the exception of Washington peak-period bus, the marginal welfare effect of increasing the subsidy is positive across modes, periods, and cities starting at subsidy levels of 50 percent. Most of these marginal welfare gains are between about 0.2 and 0.6 cents per passenger mile per one-cent increase in subsidy. Even starting at current subsidy levels, which are typically well above 50 percent, the marginal welfare effects from further lowering transit fares are positive in 9 of 12 cases.

The reasons for these results can be discerned in the figures for individual components of marginal welfare at current subsidies (in the top part of Table 3). In all cases the marginal supply cost exceeds the fare at current prices, causing an incremental welfare *loss* from this source between 0.04 and 1.27 cents per passenger mile. However, in almost all cases this loss is outweighed by incremental welfare *gains* from the combination of net scale economies and externality benefits. Washington peak-period bus is the exception here because of its especially high marginal supply cost. Welfare effects from interactions among transit modes play a reinforcing but generally more modest role, given that most of the extra passengers on transit were previously driving.

Although the contributions of net scale economies and externalities to marginal welfare vary considerably, one or the other is important in almost every case. In 7 of the 12 cases, net scale economies are substantial—between 0.3 and 1.5 cents per passenger mile. Net scale economies are larger for bus than for rail, and larger for off-peak than for peak travel; the reasons for this, already mentioned, are amplified by the greater price-responsiveness of passenger demand (and hence of service frequency or route density) in the cases of bus and of off-peak travel. Only for peak-rail service in London are scale economies fully offset by higher occupancy costs, presumably reflecting London's famous subway crowding and its already high service frequency and density. As for externalities, most of the welfare gains come from reducing road

congestion; those gains are especially large in Los Angeles and London, except for off-peak bus service.¹⁴

The lower part of Table 3 shows the optimal subsidy level, expressed as a percentage of average operating costs per passenger mile. We find that optimal fare subsidies are 68 to 90 percent or more of average operating costs in 11 of 12 cases; Washington peak bus is the exception, where the optimal subsidy is “only” 44 percent. Again, some combination of net scale economies and diversion of auto externalities explains a major part of the optimal subsidy. The gap between the average and marginal cost of expanding passenger miles, due mostly to the savings in agency costs when occupancy rather than service frequency is increased, also plays a consistently important role.

4.2. Sensitivity Analysis

To explore how robust the above results are, we first vary parameters with potential significance for marginal welfare effects and then consider different assumptions about agency adjustment.

Table 4 reports results from varying travel demand elasticities, congestion costs, the value of wait time at transit stops, the costs of increasing vehicle size, and per unit operating costs. In all cases welfare effects change in the expected direction. For example, when travel is more price responsive, the size of the welfare effect is magnified but the sign is unaffected, while lower unit operating costs lower the absolute value of the marginal cost/price gap, implying a larger overall welfare effect.¹⁵ For some perturbations welfare effects change noticeably. For example, off-peak bus is sensitive to the value of wait time through its effect on scale economies, and peak rail in Los Angeles and London is sensitive to congestion costs. Nonetheless, our basic qualitative finding—that marginal welfare effects are positive at current subsidies in the majority of cases—is unaffected. Even starting at 50 percent subsidies (results not shown), our sensitivity results show that Washington peak bus is the only case with negative marginal welfare, with two small exceptions when agency operating costs are half again as high as in our baseline.

¹⁴ In the case of Los Angeles, all components of our estimated welfare changes are compounded by existing low fares because we assume constant-elasticity demand functions, such that a unit fare change is a larger percentage of the existing fare.

¹⁵ For this simulation we adjust fares to keep subsidies constant as a proportion of operating costs.

In Table 5 we compare two scenarios: an increase in demand for passenger miles is met entirely through increased vehicle supply ($\varepsilon_V = 1.0$), versus half of the increase is met through increased vehicle miles and half through increased occupancy ($\varepsilon_V = 0.5$). Marginal welfare effects are generally lower in the former case than in our baseline, and larger in the latter. This follows because the agency costs of accommodating extra passenger miles through more vehicle miles, net of scale economies, is greater than the agency and user costs of accommodating extra passengers through higher occupancy¹⁶ (London off-peak bus is the lone exception). But again, our basic result—that subsidies of at least 50 percent of operating costs are welfare improving at the margin—is robust. The main reason is that much of the change in agency supply costs per passenger mile, as we vary ε_V , is offset by changes in net scale economies, as can be seen from the decomposition of welfare effects in the table.¹⁷

Finally, suppose that the agency does not suboptimize over route density/service frequency and vehicle size/load factor, and thus the conditions in (10) no longer hold. Suppose, for example, that service frequency is excessive relative to route density. In this case marginal wait costs will underestimate marginal access costs, and correspondingly, the marginal benefit from scale economies MB_{scale}^{ij} in (12b) will be understated by a factor of $(1 - \varepsilon_f^{ij}) \cdot \left\{ \rho^A a^{ij} \eta_a^{ij} / \rho^W w^{ij} \eta_w^{ij} - 1 \right\}$, where $1 - \varepsilon_f^{ij}$ is the fraction of marginal changes in vehicle miles that are met through increased density rather than increased service frequency. If, for illustration, $\varepsilon_f^{ij} = 0.5$ and marginal access costs exceed marginal wait costs by 50 percent, then MB_{scale}^{ij} will be understated by 25 percent. However, we have already illustrated the effect of different values for MB_{scale}^{ij} when we varied the wait cost parameter in Table 4, and for the most part, results were only moderately sensitive to different values. Similarly, relaxing suboptimization over vehicle size and crowding has essentially the same effect as varying the cost of increasing vehicle size, which again does not modify our main qualitative conclusions.

¹⁶ That is, the increase in the “marginal cost/price gap” component of MW , as ε_V is lowered from 1.0 to 0.5 in Table 5, is greater than the decrease in the “net scale economy” component.

¹⁷ The externality component also changes in some cases, especially for London off-peak bus; this is from the change in external costs from the transit vehicle itself as we alter the response of vehicle miles to passenger miles.

4.3. Relation to Recent Studies

Our findings on the marginal welfare of increasing current subsidies in London are consistent with those of Glaister (1984): he estimates that a uniform reduction in all London transit fares would produce net social benefits of £0.41 per £1 of subsidy; our estimate is 0.52.¹⁸ Our findings on optimal subsidies for London are somewhat higher than those of Glaister and Lewis (1978), who found them to be about 50 to 60 percent of operating costs in their preferred case. However, they obtain widely differing results for different parameter assumptions, and it is difficult to pinpoint the source of differences with our results.

Our results are also broadly consistent with two studies of Chicago and several Australian cities, which find that although service levels are sometimes inefficiently high, fare subsidies could generally be increased with positive effects on welfare (Savage 1997; Dodgson 1986). Our findings of very high optimal subsidies for peak-period transit are similar to those of Van Dender and Proost (2004) for Brussels, in both cases because of the high value of reducing car traffic. However, we do not find a case, as they do, for much higher peak fares; we are unsure of the reason for the difference, but we suspect that scale economies are somewhat more important in our model than theirs.

Our results contrast with those of Winston and Shirley (1998), who find optimal subsidies to be very small; however, this is not surprising because they do not include an explicit component for scale economies and they simultaneously optimize over transit prices and auto congestion tolls, effectively eliminating externality benefits from diverting drivers onto transit. Winston and Maheshri (2007) estimate that operating the Washington rail system produces a net annual aggregate welfare *loss* of \$195 million, not counting their adjustment for distortionary tax finance of agency deficits. This may appear to conflict with our finding that the current subsidy should be increased on welfare grounds. However, a closer look reconciles the apparent difference.

First, Winston and Maheshri (2007) include annualized capital cost of \$232 million as a subtraction from net benefits, which are not relevant to the question of whether operations should continue on the current system. Second, they use several assumptions that are different from ours: they do not include scale economies or interactions with bus transit, they measure unit

¹⁸ This is calculated as the average of the MW at current subsidy for the two modes and two time periods as shown in Table 3, weighted by passenger miles as shown in Table 2.

operating costs as 28 percent higher than we do, and they implicitly assume that vehicle miles change in proportion to passenger miles ($\varepsilon_V = 1$). If we run our model under these assumptions, we find optimal subsidies of 50 percent for peak rail, about the same as current subsidies, but only 21 percent for off-peak rail, well below current subsidies; thus our model would call for curtailing, while not fully eliminating, rail operations. Third, Winston and Maheshri use an own-fare demand elasticity for rail of -0.97; using our smaller assumed rail demand elasticities would roughly triple their estimated consumer surplus benefits from the rail system (by raising the measured value of service to current users), adding \$564 million in benefits. Making just this adjustment and excluding capital costs turns their estimate into a net aggregate welfare *gain* of \$600 million. This latter figure is roughly in line with the findings of Nelson et al. (2007), based on a transport network model; they estimate that the Washington rail system produces an aggregate annual welfare gain of more than \$700 million.¹⁹

5. Conclusion

Our analysis suggests that today's substantial operating subsidies for transit systems appear to be warranted on efficiency grounds, at least for the three major metropolitan areas studied. The main caveat is that this leaves aside the possibility that some of the subsidy may be lost to inefficiency or hijacked by labor unions. Thus, our analysis is most applicable to a transit agency with incentives to achieve the "best-practice" cost level; policies that might promote such cost minimization have been discussed extensively elsewhere (e.g., Wachs 1989; Winston and Shirley 1998).

Another caveat is that we have not incorporated the burden on the broader tax system from the need to finance agencies' operating deficits. To the extent that distortionary taxes elsewhere in the economy, such as income taxes, are higher than they would otherwise be, this causes efficiency losses—for example, by deterring work effort. On the other hand, lower transportation and commuting costs can stimulate more economic activity and labor force participation at the margin (Parry and Bento 2001).

Suppose, for simplicity, that the rest of the tax system is collapsed into a single tax rate on labor income. In this case the optimal externality tax or subsidy is given by the Pigouvian

¹⁹ Nelson et al. (2007) exclude scale economies, but their savings in congestion costs are larger than both ours and Winston and Maheshri's because they include parking search costs.

tax/subsidy divided by the marginal cost of public funds (Bovenberg and Goulder 2002), assuming the policy is revenue neutral and the priced good or service is an average substitute for leisure compared with other goods. The latter assumption seems a plausible approximation, at least for the bulk of transit travel, which occurs during a peak commuting period. A typical estimate for the marginal cost of public funds (which depends on the uncompensated labor supply elasticity) is around 1.1 to 1.2; making this adjustment for fiscal considerations would therefore imply some scaling back of the optimal subsidies calculated above, but not enough to overturn the basic finding that large subsidies are still warranted.

We have also ignored distributional considerations. Such concerns might raise the optimal subsidy for high-density bus service, which is heavily patronized by lower-income people, and lower it for rail service, which typically benefits wealthier riders and owners of land near transit stations. Quantifying these additional adjustments is contentious, as it brings in value judgments about appropriate distributional weights; it also runs counter to the common view that distributional concerns are most efficiently addressed through the broader tax and benefit system.

Finally, we do not explore how optimal fares might vary across different routes—for example, a route passing through bottlenecks in the central business district compared with one serving reverse commutes or intrasuburban trips. Analyzing this issue would require a more disaggregated model that accounted for substitution effects among different links in the road and transit network.

References

- APTA. 2002. *Public Transportation Fact Book*. American Public Transportation Association, Washington, DC.
- Arnott, Richard, André de Palma, and Robin Lindsey. 1993. “A Structural Model of Peak-Period Congestion: A Traffic Bottleneck with Elastic Demand.” *American Economic Review* 83: 161–79.
- Bovenberg, A. Lans, and Lawrence H. Goulder. 2002. “Environmental Taxation.” In A. Auerbach and M. Feldstein, eds., *Handbook of Public Economics* (second edition). New York: North-Holland, forthcoming.
- De Borger, Bruno, Inge Mayeres, Stef Proost, and Sandra Wouters. 1996. “Optimal Pricing of Urban Passenger Transport: A Simulation Exercise for Belgium.” *Journal of Transport Economics and Policy* 30: 31–54.

- Dodgson, J.S. 1986. "Benefits of Changes in Urban Public Transport Subsidies in the Major Australian Cities." *Economic Record*, 62: 224–35.
- Dueker, Kenneth J., Thomas J. Kimpel, and James G. Strathman. 2004. "Determinants of Bus Dwell Time." *Journal of Public Transportation* 7: 21–40.
- FTA. 2003. *National Transit Database*. Federal Transit Administration, Department of Transportation, Washington, DC. Available at www.ntdprogram.com.
- Gilbert, Christopher L., and Hossein Jalilian. 1991. "The Demand for Travel and for Travelcards on London Regional Transport." *Journal of Transport Economics and Policy* 25: 3–29.
- Glaister, Stephen. 1974. "Generalised Consumer Surplus and Public Transport Pricing." *Economic Journal* 84: 849–867.
- Glaister, Stephen. 1984. "The Allocation of Urban Public Transport Subsidy." In J. LeGrand and R. Robinson (eds.), *Privatisation and the Welfare State*, London: Allen and Unwin, 177–200.
- Glaister, Stephen and David Lewis. 1978. "An Integrated Fares Policy for Transport in London." *Journal of Public Economics* 9: 341–55.
- Goodwin, P.B., 1992. "A Review of New Demand Elasticities with Special Reference to Short and Long Run Effects of Price Changes." *Journal of Transport Economics and Policy*: 26: 155–69.
- Henderson, J. Vernon. 1977. *Economic Theory and the Cities*. New York: Academic Press.
- Jansson, Jan Owen. 1979. "Marginal Cost Pricing of Scheduled Transport Services." *Journal of Transport Economics and Policy* 13: 268–94.
- Jansson, Jan Owen. 1997. "Theory and Practice of Transport Infrastructure and Public Transport Pricing." In G. de Rus and C. Nash (eds.), *Recent Developments in Transport Economics*, Aldershot, UK: Ashgate, 74–134.
- Kenworthy, Jeff, and Felix, Laube. 2001. *The Millennium Cities Database for Sustainable Transport*. International Union of Public Transport, Brussels.
- Kraus, Marvin. 1991. "Discomfort Externalities and Marginal Cost Transit Fares." *Journal of Urban Economics* 29: 249–59.
- Lago, Armando M., Patrick D. Mayworm, and J. Matthew McEnroe. 1981. "Further Evidence on Aggregate and Disaggregate Transit Fare Elasticities." *Transportation Research Record* 799: 42–47.

- Lindberg, Gunnar. 2001. "Traffic Insurance and Accident Externality Charges." *Journal of Transport Economics and Policy* 35: 399–416.
- Mohring, H.. 1972. "Optimization and Scale Economies in Urban Bus Transportation." *American Economic Review* 62: 591–604.
- Nash, Chris. 1988. "Integration of Public Transport: an Economic Assessment." In J.S. Dodgson and N. Topham (eds.), *Bus Deregulation and Privatisation*, Aldershot, UK: Avebury, 97–118.
- Nelson, Peter, Andrew Baglino, Winston Harrington, Elena Safirova, and Abram Lipman. 2007. "Transit in Washington DC: Current Benefits and Optimal Level of Provision." *Journal of Urban Economics*, forthcoming.
- NRC. 2002. *Effectiveness and Impact of Corporate Average Fuel Economy (CAFE) Standards*. National Research Council, National Academy of Sciences, Washington, DC, National Academy Press.
- Paulley, Neil, Richard Balcombe, Roger Mackett, Helena Titheridge, John Preston, Mark Wardman, Jeremy Shires, and Peter White. 2006. "The Demand for Public Transport: The Effects of Fares, Quality of Service, Income and Car Ownership." *Transport Reviews* 13: 295–306.
- Parry, Ian W.H., and Antonio M. Bento. 2001. "Revenue Recycling and the Welfare Effects of Road Pricing." *Scandinavian Journal of Economics* 103: 645–71.
- Parry, Ian W.H., and Kenneth A. Small. 2005. "Does Britain or the United States Have the Right Gasoline Tax?" *American Economic Review* 95: 1276–89.
- Pratt, Richard H., Texas Transportation Institute, Cambridge Systematics, Parsons Brinkerhoff Quade & Douglas, SG Associates, and McCollom Management Consulting. 2000. *Traveler Response to Transportation System Changes: Interim Handbook*. Transportation Research Board, Transit Cooperative Research Program Web Document 12 (March).
- Savage, Ian. 1997. "Evaluating Transit Subsidies in Chicago." *Journal of Public Transportation*, 1: 93–117.
- Schrank, David, and Tim Lomax. 2003. *The 2003 Urban Mobility Report*, Texas Transportation Institute, Texas A&M University System, College Station, Texas.
- Small, Kenneth A. 1992. *Urban Transportation Economics*. Fundamentals of Pure and Applied Economics, Volume 51, Harwood Academic Press, Chur, Switzerland.
- Small, Kenneth A. 2004. "Road Pricing and Public Transport." In Georgina Santos, ed., *Road Pricing: Theory and Evidence*, Elsevier, 133–58.

- Small, Kenneth A., and Jose A. Gomez-Ibanez. 1999. "Urban Transportation." In Paul Cheshire and Edwin S. Mills, eds., *Handbook of Regional and Urban Economics, Volume 3: Applied Urban Economics*, Amsterdam, North-Holland, 1937–99.
- Talvitie, Antti. 1973. "A Direct Demand Model for Downtown Worktrips." *Transportation* 2: 121–52.
- Tisato, Peter. 1998. "Optimal Bus Subsidy and Cross Subsidy with a Logit Choice Model." *Journal of Transport Economics and Policy* 32: 331–50.
- TfL. 2003. *London Travel Report 2003*. London: Transport for London. www.tfl.gov.uk/tfl/reports_library_stats.shtml, accessed June 11, 2004.
- TfL. 2004a. *About London Buses*. www.tfl.gov.uk/buses/cib_about.shtml, accessed July 30, 2004.
- TfL. 2004b. *Amazing Facts*. <http://tube.tfl.gov.uk/content/faq/facts.asp>, accessed July 30, 2004.
- U.K. DfT. 2003. *Transport Statistics Great Britain*. Department for Transport, London. Available at <http://www.dft.gov.uk/pgr/statistics/datatablespublications/public/annualbulletins/abulletinofpublictransportst5264>.
- U.K. DOT, 2003. *Transport Statistics Great Britain*. London: Department for Transport.
- U.K. ONS. 2004. "Hours Worked: Growth Levels Off." In: *Labour Market in 2002*, Office for National Statistics. www.statistics.gov.uk/cc/nugget.asp?id=696, accessed June 11, 2004.
- U.S. BLS. 2004. "2002 Metropolitan Area Occupational Employment and Wage Estimates." Bureau of Labor Statistics, www.bls.gov/oes/2002/oessrcma.htm.
- U.S. DOC. 2003. *Statistical Abstract of the United States 2003*. U.S. Department of Commerce, Census Bureau, Washington, DC.
- U.S. FHWA. 1997. *1997 Federal Highway Cost Allocation Study*. U.S. Federal Highway Administration, Department of Transportation, Washington, DC. www.fhwa.dot.gov/policy/hcas/final/, accessed July 30, 2004.
- U.S. OMB. 1992. *Guidelines and Discount Rates for Benefit-Cost Analysis of Federal Programs*, Circular No. A-94, Revised, Section 8. U.S. Office of Management and Budget, Washington, October.
- U.S. TRB. 2000. *Highway Capacity Manual*. Transportation Research Board, National Research Council, Washington, DC.

Van Dender, Kurt, and Stef Proost. 2004. "Optimal Urban Transport Pricing in the Presence of Congestion, Economies of Density and Costly Public Funds." Working paper, Department of Economics, University of California at Irvine.

Viton, Philip A. 1983. "Pareto-Optimal Urban Transportation Equilibria." In Theodore E. Keeler, ed., *Research in Transportation Economics, Volume 1*. Greenwich, Connecticut: JAI Press, 75–101.

Wachs, Martin. 1989. "U.S. Transit Subsidy Policy: In Need of Reform," *Science* 244: 1545–49.

Wardman, Mark. 2001. A Review of British Evidence on Time and Service Quality Valuations. *Transportation Research E* 36: 107–28.

Winston, Clifford, and Chad Shirley. 1998. *Alternative Route: Toward Efficient Urban Transportation*. Brookings Institution, Washington, DC.

Winston, Clifford, and Vikram Maheshri. 2007. "On the Social Desirability of Urban Rail Transit Systems." *Journal of Urban Economics*, forthcoming.

Appendix A. Analytical Derivations

Equation (10): Agency optimization over route density (D) and vehicle size (n)

Combining (1), (4), and (5), the household's indirect utility function in (7) is defined by

$$(A1) \quad \tilde{U} = \tilde{u}(\{p^{ij}, t^{ij}, w^{ij}, a^{ij}, c^{ij}\}, TAX) - Z$$

$$= \underset{X, \{M^{ij}\}, \lambda}{Max} \left\{ u\left[X, M(\{M^{ij}\}), \Gamma\left(\sum_{ij} t^{ij} M^{ij}, \sum_{ij} w^{ij} M^{ij}, \sum_{ij} a^{ij} M^{ij}, \sum_{ij} c^{ij} M^{ij}\right)\right] - \sum_{ij} z^{ij} V^{ij} \right\}$$

$$+ \lambda \cdot [I - TAX - X - \sum_{ij} p^{ij} M^{ij}]$$

From the agency's point of view, (A1) can be transformed into a social utility function by substituting the various definitions and constraints of the system, namely, (2), (3), (8), and (9). In doing so, the revenues $\sum_{ij} p^{ij} M^{ij}$ in the government's budget constraint (9) cancel those in the individual's budget constraint in the last term of (A1); prices appear only insofar as M^{ij} depends on them through the consumer's demand functions. The resulting social utility function can be optimized by setting $\lambda = u_X$ (the first-order condition for X) and then by setting to zero its partial derivatives with respect to D , n , V , and either M or p . (Henceforth we omit the ij superscripts for simplicity and understand the preceding statement to apply to each i and j .) Here it is convenient to use M as the agency's choice variable; that is, we hold M constant in taking the other three derivatives. We consider two of those in this subsection, deferring the third (V) till later. Each is a partial derivative, holding the other three variables constant. Thus, in optimizing route density and vehicle size, we hold constant M and V , which implies also that occupancy $o \equiv M/V$ is constant.

Route density affects user waiting and access costs, and vehicle size affects user crowding costs and agency operating costs OC . Thus each first-order condition for optimization has two terms, and each term involves only the same i and j , so we can continue to omit the ij superscripts without ambiguity:

$$(A2) \quad 0 = \frac{\partial \tilde{U}}{\partial D} = \tilde{U}_w \frac{\partial w}{\partial D} + \tilde{U}_a \frac{\partial a}{\partial D} = -\lambda M \rho^w w_f \frac{\partial f}{\partial D} - \lambda M \rho^a a_D$$

$$(A3) \quad 0 = \frac{\partial \tilde{U}}{\partial n} = \tilde{U}_c \frac{\partial c}{\partial n} - \tilde{U}_I \frac{\partial OC}{\partial n} = -\lambda M \rho^c c_l \frac{\partial l}{\partial n} - \lambda V t \frac{dK}{dn}$$

where w_f , a_D , and c_l are derivatives of the functions defined in (3b), and we have used the definitions of ρ^k from (6b). The partial derivatives on the right-hand sides of (A2) and (A3) can be computed using definitions (3) and (8), holding V , M , and o constant. This yields $\partial f / \partial D = -f / D$, $\partial l / \partial n = -l / n$, and $dK/dn = k_2$. Inserting these and dividing each equation by λM yields (10).

Equation (11). Marginal welfare effects of reduction in peak-rail fare

Partially differentiating (A1) and applying (6b) gives

$$(A2a) \quad \tilde{U}_{p^{ij}} \equiv -\lambda M^{ij}; \quad \tilde{U}_{t^{ij}} = -\lambda \rho^T M^{ij}; \quad \tilde{U}_{w^{ij}} = -\lambda \rho^W M^{ij}; \quad \tilde{U}_{a^{ij}} = -\lambda \rho^A M^{ij}; \quad \tilde{U}_{c^{ij}} = -\lambda \rho^C M^{ij};$$

$$(A2b) \quad \tilde{U}_{TAX} = -\lambda = -u_X; \quad \tilde{U}_Z = -1$$

Totally differentiating (A1) shows that when the agency changes peak-rail price p^{PR} , utility changes according to

$$(A3) \quad \frac{d\tilde{U}}{dp^{PR}} = \tilde{U}_{p^{PR}} + \sum_{ij} \left\{ \tilde{U}_{t^{ij}} \frac{dt^{ij}}{dp^{PR}} + \tilde{U}_{w^{ij}} \frac{dw^{ij}}{dp^{PR}} + \tilde{U}_{a^{ij}} \frac{da^{ij}}{dp^{PR}} + \tilde{U}_{c^{ij}} \frac{dc^{ij}}{dp^{PR}} \right\} \\ + \tilde{U}_{TAX} \frac{dTAX}{dp^{PR}} - \sum_{ij} z^{ij} \frac{dV^{ij}}{dp^{PR}}$$

From (A2) and (A3), we obtain

$$(A4a) \quad MW^{PR} \equiv -\frac{1}{\lambda} \frac{d\tilde{U}}{dp^{PR}} = M^{PR} + \sum_{ij} M^{ij} \left\{ \rho^T \frac{dt^{ij}}{dp^{PR}} + \rho^W \frac{dw^{ij}}{dp^{PR}} + \rho^A \frac{da^{ij}}{dp^{PR}} + \rho^C \frac{dc^{ij}}{dp^{PR}} \right\} \\ + \frac{dTAX}{dp^{PR}} + \frac{1}{\lambda} \sum_{ij} z^{ij} V_M^{ij} \frac{dM^{ij}}{dp^{PR}}$$

where $V_M^{ij} \equiv dV^{ij}/dM^{ij}$ is a constant ($1/o^{iCAR}$) for $j=CAR$ and depends on the transit agency's operating policy for $j=B, R$. To keep track of its parts, we write the components of (A4a) as

$$(A4b) \quad MW^{PR} = M^{PR} + USERTIM + WAITACC + CROWD + \frac{dTAX}{dp^{PR}} + POLLACC$$

where $WAITACC$ includes the terms involving ρ^W and ρ^A and $POLLACC$, the last term in (A4a), represents changes in pollution and accident externality costs.

We can compute $dTAX/dp^{PR}$ by rearranging (9) with only TAX on the left-hand side, differentiating it, and using (2a) and (8) to get

$$(A5a) \quad \frac{dTAX}{dp^{PR}} = -M^{PR} - \sum_i \tau^{iA} V_M^{iCAR} M_{PR}^{iCAR} - \sum_i \sum_{j \neq CAR} p^{ij} M_{PR}^{ij} + \sum_i \sum_{j \neq CAR} K^{ij} \cdot \left(t^{ij} V_M^{ij} M_{PR}^{ij} + V^{ij} \frac{dt^{ij}}{dp^{PR}} \right) \\ + \sum_i \sum_{j \neq CAR} t^{ij} V^{ij} k_2^{ij} \frac{dn^{ij}}{dp^{PR}}$$

where we hold constant τ^{iA} and all transit prices other than p^{PR} . It is convenient to write the terms in (A5a) as changes in particular financial flows:

$$(A5b) \quad \frac{dTAX}{dp^{PR}} = -M^{PR} - FUELREV - TRANSITREV + (OPSUPPLY + OPCONG) + VEHSIZE$$

where the first term is changes in peak-rail revenue from existing passengers; the second is changes in fuel tax revenue; the third is changes in transit fare revenue due to mode and time-of-day shifts; the fourth is changes in transit operating cost related to travel time (divided into two parts: changes due to shifts among different modes and times of day with different average supply costs, and effects of congestion); and the last is changes in transit operating cost related to vehicle size. Note that new revenues reduce the lump-sum TAX that must be levied, whereas new costs increase it.

Substituting (A5b) into (A4b), we see that the terms M^{PR} cancel, and we can rearrange the other parts into a more convenient order for further calculation, as follows:

$$(A6) \quad MW^{PR} = (WAITACC - TRANSITREV + OPSUPPLY) \\ + (USERTIM + OPCONG + POLLACC - FUELREV) \\ + (CROWD + VEHSIZE)$$

It is useful to summarize the definitions of elasticities of bus and rail travel characteristics, recalling that all are defined so as to be positive:

$$(A7a) \quad \eta_w^{ij} = -\frac{f^{ij}}{w^{ij}} w_f^{ij}, \quad \eta_a^{ij} = -\frac{D^{ij}}{a^{ij}} a_D^{ij}, \quad \eta_c^{ij} = \frac{l^{ij}}{c^{ij}} c_l^{ij}, \quad \forall i, j \neq CAR$$

$$(A7b) \quad \varepsilon_V = \frac{M^{ij}}{V^{ij}} V_M^{ij} = o^{ij} V_M^{ij}, \quad 1 - \varepsilon_V = \frac{M^{ij}}{o^{ij}} o_M^{ij} = V^{ij} o_M^{ij}, \quad \forall i, j \neq CAR$$

We also define how service frequency and route density change with vehicle miles, and how vehicle size and load factors change with occupancy, as follows:

$$(A7c) \quad \varepsilon_f^{ij} = \frac{V^{ij}}{f^{ij}} f_V^{ij} = D^{ij} f_V^{ij}, \quad 1 - \varepsilon_f^{ij} = \frac{V^{ij}}{D^{ij}} D_V^{ij} = f^{ij} D_V^{ij}, \quad \forall i, j \neq CAR$$

$$(A7d) \quad \varepsilon_n^{ij} = \frac{o^{ij}}{n^{ij}} n_o^{ij} = l_o^{ij} n_o^{ij}, \quad 1 - \varepsilon_n^{ij} = \frac{o^{ij}}{l^{ij}} l_o^{ij} = n^{ij} l_o^{ij}, \quad \forall i, j \neq \text{CAR}$$

We now proceed to compute key derivatives in (A4a) and (A5a) in terms of $M_{PR}^{ij} \equiv dM^{ij}/dp^{PR}$. The travel time derivative can be written, using (2a), (3a), and (A7b), as

$$(A8a) \quad \frac{dt^{ij}}{dp^{PR}} = (t_{CAR}^{ij} / o^{iCAR}) M_{PR}^{iCAR} + (t_B^{ij} / o^{iB}) \varepsilon_V M_{PR}^{iB} + (\theta^j / V^{ij})(1 - \varepsilon_V) M_{PR}^{ij}$$

where $t_{CAR}^{ij} \equiv dt^j / dV_{iCAR}$ and $t_B^{ij} \equiv dt^j / dV_{iB} = \alpha_B t_{CAR}^{ij}$. Note that $t_{CAR}^{iR} = t_B^{iR} = 0$ by our assumption that rail speeds are unaffected by road traffic. Similarly, the waiting, access, and crowding derivatives in (A4), which apply only for $j \neq \text{CAR}$, can be written using (2), (3), and (A7) as

$$(A8b) \quad \frac{dw^{ij}}{dp^{PR}} = w_f^{ij} \frac{df^{ij}}{dM^{ij}} M_{PR}^{ij} = -w^{ij} \eta_w^{ij} \varepsilon_f^{ij} \varepsilon_V M_{PR}^{ij} / M^{ij}$$

$$(A8c) \quad \frac{da^{ij}}{dp^{PR}} = a_D^{ij} \frac{dD^{ij}}{dM^{ij}} M_{PR}^{ij} = -a^{ij} \eta_a^{ij} \cdot (1 - \varepsilon_f^{ij}) \varepsilon_V M_{PR}^{ij} / M^{ij}$$

$$(A8d) \quad \frac{dc^{ij}}{dp^{PR}} = c_l^{ij} \frac{dl^{ij}}{dM^{ij}} M_{PR}^{ij} = c^{ij} \eta_c^{ij} (1 - \varepsilon_n^{ij})(1 - \varepsilon_V) M_{PR}^{ij} / M^{ij}$$

We now examine the terms in (A6) in groups. We begin by using (A8b) and (A8c) to compute *WAITACC* as given in (A4), using (10a) to simplify:

$$(A9) \quad \begin{aligned} \text{WAITACC} &= \sum_i \sum_{j \neq \text{CAR}} \left(\rho^W \frac{dw^{ij}}{dp^{PR}} + \rho^A \frac{da^{ij}}{dp^{PR}} \right) M^{ij} = - \sum_i \sum_{j \neq \text{CAR}} \rho^W w^{ij} \eta_w^{ij} \varepsilon_V M_{PR}^{ij} \\ &= - \sum_i \sum_{j \neq \text{CAR}} MB_{scale}^{ij} M_{PR}^{ij} \end{aligned}$$

where the last equality applies definition (12b). This accounts for all the terms in (11) involving MB_{scale} . As for the other terms in the first group in (A6), we note that *TRANSITREV*, the third term in (A5a), accounts for all the terms in (11) involving p . We also see that *OPSUPPLY*, as defined by the first of the two terms involving K^{ij} in (A5a), can be written using (A7b) as

$$(A10) \quad \text{OPSUPPLY} = \sum_i \sum_{j \neq \text{CAR}} (\varepsilon_V / o^{ij}) K^{ij} t^{ij} M_{PR}^{ij} = \sum_i \sum_{j \neq \text{CAR}} MC_{supply}^{ij} M_{PR}^{ij}$$

where the last equality uses definition (12a). Thus *OPSUPPLY* accounts for all the terms in (11) involving MC_{supply} .

We now turn to the second group of terms in (A6). The terms *USERTIM* and *OPCONG*, which are the terms in (A4a) and (A5a) involving dt^{ij}/dp^{PR} , can be combined and written, using (A8a), as

$$\begin{aligned}
 USERTIM + OPCONG &\equiv \sum_{ij} (M^{ij} \rho^T + K^{ij} V^{ij}) \frac{dt^{ij}}{dp^{PR}} \\
 &= \sum_i \sum_{j=CAR,B} (\rho^T M^{ij} + K^{ij} V^{ij}) (t_{CAR}^{ij} / o^{iCAR}) M_{PR}^{iCAR} \\
 &\quad + \sum_i \sum_{j=CAR,B} \varepsilon_V \cdot (\rho^T M^{ij} + K^{ij} V^{ij}) (t_B^{ij} / o^{iB}) M_{PR}^{iB} \\
 &\quad + \sum_i \sum_{j=B,R} (1 - \varepsilon_V) (\rho^T M^{ij} + K^{ij} V^{ij}) (\theta^j / V^{ij}) M_{PR}^{ij}
 \end{aligned}$$

where we have adopted the notational convention that $K^{iCAR}=0$. Using the fact that $t_B^{ij} = \alpha_B t_{CAR}^{ij}$, the definition $o^{ij}=M^{ij}/V^{ij}$, and definitions (12d), we obtain

$$\begin{aligned}
 (A11) \quad USERTIM + OPCONG &= \sum_i (MC_{cong}^{iCAR} / o^{iCAR}) M_{PR}^{iCAR} + \sum_i \varepsilon_V \cdot (MC_{cong}^{iB} / o^{iB}) M_{PR}^{iB} \\
 &\quad + \sum_i \sum_{j=B,R} (1 - \varepsilon_V) MC_{dwell}^{ij} M_{PR}^{ij}
 \end{aligned}$$

These terms are components of sums of MC_{ext}^{ij} as defined in (12c). Next we obtain some other components of those same sums. Using the definition of ε_V and the fact that $\lambda=u_X$, the change in external costs of pollution and accidents is

$$(A12) \quad POLLACC \equiv \frac{1}{\lambda} \sum_{ij} z^{ij} V_M^{ij} M_{PR}^{ij} = \sum_i \frac{z^{iCAR}}{u_X} \cdot \frac{1}{o^{iCAR}} \cdot M_{PR}^{iCAR} + \sum_i \sum_{j \neq CAR} (\varepsilon_V / o^{ij}) \cdot \frac{z^{ij}}{u_X} \cdot M_{PR}^{ij}$$

Finally, the fuel tax revenue term in (A5) is

$$(A13) \quad -FUELREV = \sum_I (-\tau^{iCAR} / o^{iCAR}) M_{PR}^{iCAR}$$

Adding equations (A11)-(A13) and applying definitions (12c) yields

$$\sum_i MC_{ext}^{iCAR} M_{PR}^{ij} + \sum_i \sum_{j=B,R} MC_{ext}^{ij} M_{PR}^{ij}$$

which accounts for all the terms in (11) involving MC_{ext} .

Finally, we consider the last group of terms in (A6), involving crowding and the costs undertaken to avoid it. Using (A7b), A7d), (A8d), and (10b), these terms add to

$$\begin{aligned}
CROWD + VEHSIZE &= \sum_i \sum_{j \neq CAR} (1 - \varepsilon_v)(1 - \varepsilon_n^{ij}) \rho^C c^{ij} \eta_c^{ij} M_{PR}^{ij} + \sum_i \sum_{j \neq CAR} t^{ij} V^{ij} k_2^{ij} n_o^{ij} o_M^{ij} M_{PR}^{ij} \\
&= \sum_i \sum_{j \neq CAR} (1 - \varepsilon_v)(1 - \varepsilon_n^{ij}) (t^{ij} k_2^{ij} n^{ij} / o^{ij}) M_{PR}^{ij} + \sum_i \sum_{j \neq CAR} (1 - \varepsilon_v) \varepsilon_n^{ij} (t^{ij} k_2^{ij} n^{ij} / o^{ij}) M_{PR}^{ij} \\
&= \sum_i \sum_{j \neq CAR} MC_{occ}^{ij} M_{PR}^{ij}
\end{aligned}$$

which accounts for the terms in (11) involving MC_{occ} . We have now accounted for all terms in (11), which completes the proof.

Transit agency optimization over vehicle miles of service (V).

Now consider optimizing with respect to V , which we consider only for some scenarios (including the baseline scenario). We compute the optimal value of ε_v under certain additional simplifying assumptions, namely these three:

- Elasticities of waiting and access times (defined positively) are all equal to a common value ($\eta_w^{ij} = \eta_a^{ij} \equiv \zeta$);
- The transit agency ignores its own vehicles' contributions to congestion ($t_B^{ij} = 0$) and to other externalities ($z^{iB} = z^{iR} = 0$);
- Dwell time for entering and exiting passengers is negligible ($\theta^B = \theta^R = 0$).

The first bullet is an assumption common to the simpler models of Mohring effects—for example, that of Small (2004). A special case is when average waiting time is half the interval between vehicles, and average access time is proportional to the distance between parallel transit lines; then $\zeta = 1$.

Those assumptions enable us to derive a simple condition for maximizing (A1) with respect to the agency's choice variables, for given travel demands $\{M^{ij}\}$. In what follows, we suppress superscripts for simplicity. Maximizing with respect to D and n again yields (10). Given

our first simplifying assumption, we see immediately from (10a) that average waiting cost and access cost are equated:

$$(A14) \quad \rho^W w = \rho^A a$$

This result is also in Jansson (1997). Since $D=V/f$, it can be written as

$$(A15) \quad \rho^W \alpha_w f^{-\zeta} = \rho^A \alpha_a \cdot (V / f)^{-\zeta}$$

where we have substituted in the constant-elasticity functions $w = \alpha_w f^{-\zeta}$ and $a = \alpha_a D^{-\zeta}$ describing waiting and access times, respectively. Solving (A15) for f , we see that it is proportional to the square root of V . That is, f is adjusted when V changes with elasticity $\varepsilon_V = 1/2$. Therefore,

$$(A16) \quad \varepsilon_M = \varepsilon_V \varepsilon_V = 1/2 \varepsilon_V$$

We now consider maximizing with respect to V . Given our second assumption, V affects (A1) only through the terms involving waiting time w , access a , crowding c , and operator cost OC , the latter entering through budget constraint (9). The first-order condition is therefore

$$\begin{aligned} 0 &= \frac{\partial \tilde{U}}{\partial V} = \tilde{U}_w \frac{\partial w}{\partial V} + \tilde{U}_a \frac{\partial a}{\partial V} + \tilde{U}_c \frac{\partial c}{\partial V} + \tilde{U}_{TAX} \frac{\partial OC}{\partial V} \\ &= -\lambda M \rho^W w_f f_V - \lambda M \rho^A a_D D_V - \lambda M \rho^C c_l l_o \cdot (do / dV) - \lambda K t - \lambda t V k_2 n_o \cdot (do / dV) \end{aligned}$$

where the last equality uses the definitions in (6b) and (8) and the result $\lambda = u^X$. Dividing by λ and using (A7), (2a), and (10), this implies

$$\begin{aligned} 0 &= \rho^W w \zeta \varepsilon_f M / V + \rho^A a \zeta \cdot (1 - \varepsilon_f) o + \rho^C c \eta_c \cdot (1 - \varepsilon_n) o - k_1 t - k_2 t n \cdot (1 - \varepsilon_n) \\ &= \rho^W w \zeta M / V - k_1 t \end{aligned}$$

or

$$(A17) \quad \frac{wM}{V} = \frac{k_1 t}{\rho^W \zeta}$$

Under our second assumption, the right-hand side of (A17) is a constant as far as the agency is concerned. On the left-hand side, $w = \alpha_w f^{-\zeta}$. Therefore,

$$(A18) \quad f^{-\zeta} M V^{-1} = \text{constant}$$

Now let M change parametrically, with all the service variables f , n , and V changing in response. Differentiating the logarithm of (A18) with respect to $\log(M)$ yields

$$(A19) \quad -\zeta \varepsilon_f + 1 - \varepsilon_V = 0$$

Substituting (A16) into (A19) and solving yields $\varepsilon_V = 2/(2+\zeta)$. For the common case $\zeta=1$, this yields $\varepsilon_V = 2/3$, as in Small (2004) and a special case of Nash (1988).

The intuition for this result is somewhat subtle. If ζ is near zero, wait and access costs are relatively unaffected by vehicle miles of service, so vehicles are operated only as necessary to handle the passenger loads; thus increased passenger loads require a proportional increase in vehicle miles, *i.e.* $\varepsilon_V = 1$. If ζ is large, the operator accounts for the substantial effects on user costs by running extra vehicles for passengers' convenience even when M is small; in that case, when M increases, the operator can absorb some of the increase through higher occupancy, thereby reaping more of the advantages of scale; this means choosing a smaller value of ε_V . We take $\zeta=1$ as our base case ($\varepsilon_V = 2/3$) and consider sensitivity $\zeta \in [0, 2]$ by treating $\varepsilon_V = 1$ and $\varepsilon_V = 1/2$.

Appendix B. Assessment of Parameter Values

Here we describe our methodology for estimating parameter values along with data sources; Table 2, which is discussed in the text, summarizes our key estimates. For some parameters, breakdowns by mode or time of day are unavailable from statistical sources; in these cases we use various estimation procedures or our own judgment. U.K. monetary numbers are converted to U.S. dollars using the average 1998–2003 exchange rate of £1.0 = US\$1.6.

System aggregates. Basic data are compiled from the operating agencies and various national statistics.²⁰ For London, we allocate total passenger miles across time of day using the observed fraction of passenger trips occurring at peak period, 0.62 for rail and 0.48 for bus, and an assumed average trip length in the peak equal to 1.6 times that in the off-peak.²¹ Passenger miles per hour are then computed assuming that the peak period covers 6 hours per workday (30 hours per week) and the off-peak covers 10 hours every day (70 hours per week). We assume peak shares are each 0.05 higher for Washington (which has a high proportion of government employment) and 0.05 lower for Los Angeles (which has a smaller discrepancy between peak and off-peak vehicles per hour).²² To obtain vehicle miles per hour, we assume that observed total vehicle miles are allocated in proportion to their respective passenger miles per hour to the power $\varepsilon_V=0.67$, our baseline assumption as discussed in Appendix A.²³

For Washington and Los Angeles, automobile vehicle miles by time of day are from Shrank and Lomax (2003), and occupancy is from the 2001 National Household Travel Survey on average occupancy per trip in large metropolitan areas. For London, auto passenger and vehicle miles by time of day are from TfL (2003, Tables 1.2, 3.1 and 3.6), assuming the same ratio of peak to off-peak occupancy as for Washington.

²⁰ For the United States, see the National Transit Database (FTA 2003), and for the United Kingdom, see TfL (2003, Tables 1.1, 1.2, 3.6), TfL (2004a, b), and U.K. DfT (2003, Tables 5.3, 5.16). Rail data encompass subways and light rail but not commuter rail.

²¹ For the entire United Kingdom, commuting trips have around twice the length of trips for education, shopping, or other personal business (U.K. DOT, 2003, Table 10); however, we expect a smaller discrepancy for transit trips because of the high fixed time cost of accessing transit.

²² The Washington adjustments are in line with unpublished statistics we obtained from transit agency representatives; the Los Angeles transit authority has no such data on trips by time of day.

²³ Total vehicle miles for rail were obtained by multiplying vehicle-car miles by average cars per train; for peak periods the latter is calculated by the ratio of rail cars to trains. Off-peak train length is assumed to be slightly lower based on common observation.

Operating costs and fares. We assume that vehicle capital costs are proportional to capacity n , whereas other operating costs are independent of n . Thus in aggregate, vehicle capital costs constitute $k_1 tV$ and other operating costs $k_2 n tV$, using (8). Operating costs, aside from vehicle capital costs, are taken from the operating statistics of the transit agencies. For rail, we assume that 10 percent of these are the fixed cost of maintaining stations (F^{ir} in (8a)). When expressed per vehicle hour of service, we assume that the rest of these costs are 25 percent greater during peak than off-peak periods because of difficulties in scheduling labor for split shifts; hence we obtain k_1^{ij} in (8b).

As for vehicle capital costs, we estimate them ourselves by annualizing the purchase cost of a rail or bus car, assuming lifetimes of 25 and 12 years, respectively, and a real interest rate of 7 percent.²⁴ We allocate vehicle capital costs entirely to the peak period, on the assumption that any increase in vehicle miles in that period requires purchasing more vehicles, whereas an increase in the off-peak period does not; hence we obtain $k_2^{pj} n^{pj}$ and $k_2^{oj} = 0$ in (8b). Vehicle capital costs are 27 to 52 percent of other peak variable operating costs. Thus our assumption that they are the portion of costs that is proportional to n leads to results consistent with several other studies of size-related costs, as reviewed by Small (2004, 156 and note 13).

Fares were obtained by dividing agency passenger fare revenue by passenger mileage (for Washington rail, peak fares were higher than off-peak in 2002, but the discrepancy was modest and we ignore it).

Wait costs. Based on evidence summarized in Small (1992, 44), we assume the value per minute of in-vehicle time ρ^T is 0.5 and 0.375 times the median gross wage rate for peak and off-peak periods, respectively; hourly wages are taken to be \$16.93, \$14.19, and \$12.06 for Washington, Los Angeles, and London, respectively, and then expressed per minute.²⁵ The standard consensus has been that the value of waiting time at transit stops, ρ^W , is two to three times ρ^T ; but a meta-analysis by Wardman (2001, 109, Table 2, first column) on British studies produces a ratio

²⁴ We use U.S. nationwide figures for all vehicle prices (from APTA 2002, Table 60) except for Los Angeles rail, for which figures were available from www.mta.net/press/pressroom/facts_glance (where necessary, figures are updated to 2002 using the CPI for Transportation Equipment). The vehicle lifetimes chosen are common in the transit cost literature, and the interest rate is that recommended for cost-benefit analysis by U.S. OMB (1992).

²⁵ From U.S. BLS (2002), TfL (2003, p. 49) and U.K. ONS (2004).

between 1.47 and 1.81. To be conservative, we set $\rho^W = 1.6\rho^T$.

We obtain initial wait times and the wait time elasticity as follows. Let H be average minutes between transit vehicles at a given stop, or headway (as calculated by the inverse of vehicles per hour). When H is small, it is reasonable to assume that travelers arrive randomly at a stop and incur expected wait time $H/2$. When headways are larger, at least some travelers will use transit timetables, which, following Tisato (1998), we assume involves three time costs. The first two are planning and precautionary time required because the exact vehicle arrival time is uncertain; we assume these are 1 and 5 minutes, respectively, and each is valued at rate ρ^W . The third is the expected cost of early arrival at the destination, assuming the traveler chooses a transit vehicle arriving prior to the desired time to ensure against late arrival. This is $\rho^E H/2$ where ρ^E is the per minute cost of early arrival, assumed to be $0.2\rho^W$ (Arnott et al. 1993); that is, a minute of early arrival is equivalent to 0.2 minutes of extra planning or precautionary time. All these costs are therefore accounted for by setting $w = 6 + 0.2H/2$ for those using a timetable.

When $H < 15$, all users will arrive randomly, whereas when $H > 15$, the average wait time per trip is $(1 - \lambda) \cdot H/2 + \lambda \cdot (6 + H/10)$, where λ is the fraction of travelers following a schedule, which we assume rises linearly from zero at $H = 15$ to one at $H = 45$. The elasticity η_w^{ij} is then -1 for $H < 15$ and declines in magnitude to -0.5 at $H = 45$. Wait time, given by the above expression for headway, is divided by trip length to express it on a per mile basis, and then multiplied by ρ^W to give the initial wait cost.

Marginal benefits from scale economies and marginal cost from occupancy. These are easily computed from (12b), using above values for parameters ε_V , ρ^W , w^{ij} , η_w^{ij} , and $k_2^{ij} n^{ij}$.

Marginal congestion costs. For automobiles, MC_{cong}^{iCAR} is obtained by multiplying estimates of average delay per passenger mile by 3.7 to convert them into marginal hours of delay, and then multiplying by the value of time, ρ^T .²⁶ Average delay for the U.S. cities is obtained from dividing total person hours of delay from Schrank and Lomax (2003) by passenger miles, and allocating 85 percent of it to the peak period (this yields an average peak delay of 0.33 minutes

²⁶ This is based on averaging over relationships fit by Small (1992, 70–71) which suggest that total delay is well approximated by a power function of traffic volume, with power 4.1 in Toronto and 3.3 in Boston.

per passenger mile for Washington and 0.49 for Los Angeles). Our data provide direct estimates of average traffic speeds in Greater London during the peak and daytime off-peak periods; we add 10 percent to the latter to account for evenings and nights. Average delay is then inferred assuming a free-flow speed of 30 miles per hour. We assume the passenger-car equivalent for buses, α_B in (12d), is 4.0 for the U.S. cities and 5.0 for London, where buses are larger and cars are smaller.²⁷

Pollution and accident externalities. We start with nationwide average values from the assessment by Parry and Small (2005) of U.S. and U.K. automobile externalities: namely, 2.0 cents per vehicle mile for local pollution; 6 cents per gallon of gasoline for global warming; and 3.0 and 2.4 cents per vehicle mile for accidents in the United States and the United Kingdom, respectively. To account for greater population exposure in urban areas, we double the local pollution figure for Washington and London, and we triple it for Los Angeles, whose topography causes pollutants to disperse especially slowly. We do not adjust external accident costs because the evidence suggests that, despite higher traffic densities in urban areas, external accident risks are not necessarily, given the counteracting effect of slower-moving traffic (Lindberg 2001, 406–407).

Also from Parry and Small (2005), we assume fuel taxes of 40 cents per gallon for the U.S. cities²⁸ and 280 cents per gallon in London. We use their nationwide average fuel efficiencies of 20 and 30 miles per gallon for the off-peak period (on the assumption that most travel nationwide is in conditions similar to off-peak travel in these very large metropolitan areas) but reduce them by 25 percent in the peak period to adjust for the effect of congestion on fuel economy.

For bus, accidents costs per vehicle mile are taken to be the same as for auto because buses move more slowly and are driven by professionals, offsetting their much greater weight, but pollution is taken to be triple that for automobiles.²⁹ When expressed per passenger mile,

²⁷ U.S. FHWA (1997, Table V-23) gives the passenger-car equivalent as only 2.0; however, this is only for federal urban highways where buses stop very infrequently, and it excludes mileage on city and suburban streets.

²⁸ The federal tax was 18.4 cents per gallon; state-level taxes in California, the District of Columbia, Virginia, and Maryland were approximately 20 cents per gallon (U.S. DOC 2003, Table 730).

²⁹ These assumptions are consistent with estimates of relative external costs per vehicle mile for heavy trucks and autos in U.S. FHWA (1997, Table 13); separate estimates for bus are not available.

these external costs are small. Pollution and accident external costs per passenger mile for rail are taken to be zero.

Dwell times. For bus, we adopt the midrange values for typical boarding and alighting times from U.S. TRB (2000, Exhibit 27-9), assuming two doors for alighting and boarding. We assume cash payment for the U.S. cities and prepayment (which allows rear-door boarding) for London. This yields values of 4.275 seconds for the U.S. cities and 3.375 sec for London (for comparison, Dueker et al. 2004 estimate 5.18 sec in Portland). For rail, we use the estimate by Kraus (1991, 256) from observations in Boston, which is $1.0/N_T$ sec where N_T is the number of cars per train. In each case we divide by trip length to specify parameter θ^j . The marginal cost of increased dwell time is then calculated from (12d), using parameters already described.

Generalized price of travel. The components of q^{ij} are given by (10c); besides parameter values already described, we need the time per mile of transit vehicles t^{ij} and access and crowding elasticities η_a^{ij} and η_c^{ij} . (This is in fact the only place where we need an empirical estimate of η_c^{ij} .)

To calculate t^{ij} , we divide total vehicle miles by vehicle hours to give average speeds, over the day, of 23 and 11 miles per hour for Washington rail and bus, and 23 and 12 miles per hour for Los Angeles rail and bus. For London, we have a direct estimate of rail speed from the agency, of 20 miles per hour. For all three cities we assume the ratio of peak to off-peak speed is 1.0 for rail, while for bus it is the same as that for autos: approximately 0.8 for Washington and London and 0.75 for Los Angeles.

The access-time elasticity η_a^{ij} depends on route density in a manner similar to how the wait-time elasticity depends on service frequency. It is one if people live at uniformly distributed locations and walk to the nearest transit stop, and smaller if people living farther away choose a faster access mode with a fixed cost (e.g., park and ride). The less dense the transit network, the more important these other access modes, so the lower the elasticity. We assume other access modes have minor importance in London but more in Washington and more still in Los Angeles, and so choose $\eta_a^{ij} = 0.8, 0.65$ and 0.5 for these cities, respectively.

There is little empirical basis for gauging η_c^{ij} , which is positive only for peak service; we set it to 1.5 in the baseline, though our results are not sensitive to different assumptions (because crowding costs are relatively small).

Own-price travel demand elasticities. Our model calls for elasticities of each mode's passenger demand with respect to its own generalized price q^{ij} , denoted as η_q^{ij} . However, most empirical evidence is based on elasticities with respect to fare p^{ij} , which we denote as η_p^{ij} . We first review the evidence on η_p^{ij} , then describe how we convert to η_q^{ij} .

Based on Lago et al. (1981), Goodwin (1992), and Pratt et al. (2000), we assume that the own-fare demand elasticity, averaged over peak and off-peak time periods, is -0.5 for bus and -0.3 for rail,³⁰ and that in each case the elasticity in the off-peak period is twice that in the peak. Given that about 70 percent of passenger mileage occurs during the peak period, the values just stated imply own-fare elasticities η_p^{ij} of approximately -0.40 and -0.8 for peak and off-peak bus, and -0.24 and -0.48 for peak and off-peak rail, respectively. To convert these to generalized-price elasticities η_q^{ij} , we assume that the empirical measurement of η_p^{ij} incorporates the effects of p^{ij} on w^{ij} in (10c), as discussed in the derivation of (14c); that is, we assume

$$(B1) \quad \eta_p^{ij} = \frac{p^{ij}}{M^{ij}} \frac{dM^{ij}}{dq^{PR}} \frac{dq^{ij}}{dp^{ij}} = \eta_q^{ij} \frac{p^{ij}}{q^{ij}} \frac{dq^{ij}}{dp^{ij}}$$

where the ratio and the derivative on the right-hand side are both obtained from (10c). Thus we simply invert equation (B1) to obtain our estimates of η_q^{ij} , which we assume to be constants.

Modal diversion ratios, m_{kl}^{ij} . Pratt et al. (2000, 12–41 ff.) provide several estimates for U.S. cities of the proportion of incremental transit trips that are diverted to or from other modes following a change in transit price; typical numbers, averaged across time of day, are about 65 percent and 80 percent for Atlanta and Los Angeles, respectively. Nearly all of these shifts are to or from

³⁰ A recent review of mostly U.K. studies by Paulley et al. (2006) produces somewhat larger long-run elasticities, which they suggest is because elasticities have risen in magnitude and are higher in the United Kingdom than in other nations. Many of the studies relied upon by Paulley et al. are unpublished, and we do not feel the evidence is strong enough to apply these higher elasticities to our U.K. simulations.

cars. We assume that Washington is like Atlanta, and that peak values m_{PCAR}^{Pj} are 0.05 higher, and off-peak values m_{OCAR}^{Oj} 0.05 lower, than these average values.

Now consider the cross-elasticities between bus and rail transit. The few studies available typically find them to be about half the direct elasticities in cities with good coverage by both bus and rail transit systems, such as London and Chicago (Gilbert and Jalilian 1991, Table 3b; Talvitie 1973). Assuming equal travel volume by mode, this would imply $m_{iR}^{iB} = m_{iB}^{iR} \approx 0.5$ for $i=P,O$. However, we expect the substitutability between modes to decrease as one expands beyond the city to the metropolitan area, and to decrease more for cities with less and less well developed rail networks. We also expect them to have declined considerably from the 1970s or 1980s to the year 2000 because of increasing competition from the automobile. Finally, in the newer U.S. transit systems the bus lines are typically reconfigured to serve as feeders to the rail system, with competitive routes discontinued. Therefore, we assume the cross-mode diversion ratios to be just 10 percent for Washington ($m_{iR}^{iB} = m_{iB}^{iR} = 0.1$) and 5 percent for Los Angeles ($m_{iR}^{iB} = m_{iB}^{iR} = 0.05$).

For London, we expect less diversion to automobile and more to the other transit mode because of the smaller initial share of automobiles and travelers' more extensive transit choices. We therefore set London's diversion ratios to be like those for Washington, except 0.20 smaller for auto in the same time period, and 0.20 larger for other transit in the same time period.

Little information is available about shifts of transit riders across time periods. We assume that in each case, 10 percent of the change in transit ridership represents such shifts, and that the shifts occur entirely to the same transit mode.

Those assumptions lead to the values shown in Table 2. The fraction of extra transit trips from increased travel demand is a residual, equal to between zero and 20 percent. The review by Pratt et al. (2000) suggests that 10 percent and 26 percent of new transit trips in Los Angeles and Atlanta, respectively, represented some combination of changes in walking, trip frequency, and destination during the 1990s; given the likely further decline in this fraction due to metropolitan decentralization, this evidence is roughly consistent with our assumed values.

Table 1. Passenger Fare Subsidies for the 20 Largest U.S. Transit Authorities

	Fare subsidy, % of average operating cost per passenger mile			Passenger miles		
	rail	bus	combined	total, million	% rail	%bus
MTA New York City Transit, Brooklyn, NY	29	59	59	9,451	83	17
New Jersey Transit Corporation, Newark, NJ	50	57	47	2,548	64	36
MTA Long Island Rail Road/Bus, Jamaica, NY	53	61	47	2,147	93	7
Metro-North Commuter Railroad Co., New York, NY	40	n/a	60	2,059	100	0
Washington Metrop. Area Transit Authority, Washington, DC	40	75	45	1,899	76	24
Massachusetts Bay Transportation Authority, Boston, MA	57	79	36	1,838	82	18
Los Angeles County Metrop. Transp. Authority, Los Angeles, CA	78	71	28	1,818	21	79
Chicago Transit Authority, Chicago, IL	59	64	38	1,814	58	42
Northeast Illinois Regional Commuter Railroad Corp., Chicago, IL	56	n/a	45	1,506	100	0
Southeastern Pennsylvania Transp. Authority, Philadelphia, PA	50	62	43	1,354	65	35
San Francisco Bay Area Rapid Transit District, Oakland, CA	42	n/a	58	1,148	100	0
Metropolitan Atlanta Rapid Transit Authority, Atlanta, GA	67	71	31	722	68	32
Maryland Transit Administration, Baltimore, MD	72	74	27	631	47	53
King County Dept. of Transp. - Metro Transit Division, Seattle, WA	85	82	18	433	0	100
Metrop. Transit Authority of Harris County, Texas, Houston, TX	n/a	81	19	417	0	100
Tri-County Metrop. Transp. District of Oregon, Portland, OR	35	89	24	407	42	58
Miami-Dade Transit, Miami, FL	85	75	23	389	28	72
Dallas Area Rapid Transit, Dallas, TX	89	88	12	385	36	64
Denver Regional Transportation District, Denver, CO	63	80	21	371	12	88
Port Authority of Allegheny County, Pittsburgh, PA	82	74	26	305	10	90
Average (unweighted)	61	77	35	31,617	54	46
Average (weighted by passenger miles)	44	69	49	31,617	72	28

Source: FTA (2003).

Table 2. Selected Baseline Parameter Values

	Washington				Los Angeles				London			
	Rail		Bus		Rail		Bus		Rail		Bus	
	Peak	Off-Peak	Peak	Off-Peak	Peak	Off-Peak	Peak	Off-Peak	Peak	Off-Peak	Peak	Off-Peak
TRANSIT												
Annual passenger miles, millions	1,100	339	290	161	267	126	799	662	3,302	1,265	2,115	1,432
Passenger miles per hour, thousands	705	93	186	44	171	35	512	182	2,117	348	1,356	393
Vehicle miles per hour, thousands	4.67	1.08	13.51	4.63	1.46	0.45	27.80	12.46	15.99	4.29	82.43	32.30
Vehicle occupancy	151	86	14	10	117	77	18	15	132	81	16	12
Average operating cost, \$/ veh-mi	58	35	14	7	52	33	13	6	104	67	8	4
Avg operating cost, c/pass-mi	38	41	103	74	45	42	72	42	78	83	49	30
Marginal supply cost, c/pass-mi	23	25	69	50	27	26	48	28	47	50	33	20
Fare, c/pass-mi	20	20	20	20	8	8	14	14	25	25	20	20
Subsidy, % of average operating cost	48.6	51.9	80.7	73.1	82.6	81.6	79.9	65.5	68.2	69.9	60.1	34.4
Cost of in-vehicle travel time, c/pass-mi	43	32	93	58	37	27	79	44	35	26	69	43
Wait cost, c/pass-mi	10	23	42	72	8	19	28	47	5	16	20	59
Wait time elasticity	1.00	1.00	0.96	0.65	1.00	1.00	0.94	0.65	1.00	1.00	1.00	0.87
Marginal scale economy, c/pass-mi	6	16	27	31	5	13	17	20	3	11	13	34
Marginal cost of occupancy, c/pass-mi	3	0	9	0	3	0	5	0	5	0	5	0
Marginal external cost, c/pass-mi	1	1	9	5	1	1	9	5	1	1	32	18
Marg. congestion cost, c/pass-mi	0	0	5	1	0	0	5	1	0	0	29	15
Pollution. & accident, c/pass-mi	0	0	1	1	0	0	1	1	0	0	1	1
Marginal dwell cost, c/pass-mi	1	1	3	3	1	1	2	2	1	1	2	2
Generalized price, c/pass-mi	93	110	235	222	75	94	184	166	81	87	144	186
Elasticity of passenger demand wrt fare	-0.24	-0.48	-0.40	-0.80	-0.24	-0.48	-0.40	-0.80	-0.24	-0.48	-0.40	-0.80
Fraction of increased transit coming from												
auto--same period	0.70	0.60	0.70	0.60	0.85	0.75	0.85	0.75	0.50	0.40	0.50	0.40
same transit mode--other period	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
other transit mode--same period	0.10	0.10	0.10	0.10	0.05	0.05	0.05	0.05	0.30	0.30	0.30	0.30
increased overall travel demand	0.10	0.20	0.10	0.20	0.00	0.10	0.00	0.10	0.10	0.20	0.10	0.20
AUTO												
	Peak	Off-Peak			Peak	Off-Peak			Peak	Off-Peak		
Annual passenger-miles, millions	19,583	22,055			69,519	75,226			13,397	15,859		
Occupancy	1.34	1.45			1.34	1.45			1.40	1.51		
Marginal external cost, c/pass-mi	25	6			31	9			99	35		
Marg. congestion cost, c/pass-mi	21	2			26	3			103	37		
Pol. & acc. less fuel tax, c/pass-mi	4	4			5	5			-4	-2		

Source: See Appendix B.

Table 3. Baseline Welfare and Optimal Subsidy Estimates

	Washington				Los Angeles				London			
	Rail		Bus		Rail		Bus		Rail		Bus	
	Peak	Off-Peak	Peak	Off-Peak	Peak	Off-Peak	Peak	Off-Peak	Peak	Off-Peak	Peak	Off-Peak
Current subsidy, % of op. cost	48.6	51.9	80.7	73.1	82.6	81.6	79.9	65.5	68.2	69.9	60.1	34.4
MW at current subsidy ^a	0.24	0.33	-0.46	0.17	0.35	0.14	-0.12	0.52	0.36	-0.02	0.36	1.59
marginal cost/price gap	-0.04	-0.12	-0.99	-1.27	-0.60	-1.11	-0.92	-0.83	-0.21	-0.48	-0.27	-0.06
net scale economy	0.04	0.38	0.37	1.37	0.07	0.81	0.32	1.08	-0.01	0.21	0.16	1.48
externality	0.20	0.07	0.15	-0.04	0.79	0.34	0.46	0.10	0.47	0.25	0.35	-0.19
other transit	0.04	0.01	0.01	0.11	0.10	0.10	0.02	0.17	0.12	-0.05	0.12	0.35
MW at 50% subsidy ^a	0.23	0.34	-0.03	0.34	0.35	0.28	0.10	0.61	0.35	0.13	0.37	2.18
Optimum subsidy, % of op. cost	>90.0	88.0	44.0	82.0	>90.0	89.0	70.0	>90.0	>90.0	68.0	>90.0	>90.0
proportion of subsidy due to												
average-marginal cost gap	0.43	0.54	0.28	0.37	0.38	0.49	0.37	0.35	0.36	0.60	0.40	0.27
net scale economy	0.08	0.37	0.53	0.60	0.04	0.32	0.27	0.52	-0.02	0.20	0.14	0.64
externality	0.42	0.08	0.17	-0.02	0.51	0.14	0.35	0.05	0.52	0.25	0.34	-0.09
other transit	0.08	0.01	0.01	0.05	0.07	0.04	0.01	0.08	0.14	-0.05	0.12	0.18

Note

^a This is the marginal welfare gain from a one-cent-per-mile reduction in the passenger fare, expressed in cents per passenger mile.

Table 4. Results with Alternative Parameter Values: Marginal Welfare Effects at Current Subsidies

	Washington				Los Angeles				London			
	Rail		Bus		Rail		Bus		Rail		Bus	
	Peak	Off-Peak	Peak	Off-Peak	Peak	Off-Peak	Peak	Off-Peak	Peak	Off-Peak	Peak	Off-Peak
Baseline results	0.24	0.33	-0.46	0.17	0.35	0.14	-0.12	0.52	0.36	-0.02	0.36	1.59
Travel demand elasticities												
Increased by 30%	0.36	0.51	-0.45	0.57	0.50	0.24	-0.24	1.17	0.48	-0.03	0.48	5.37
Reduced by 30%	0.18	0.20	-0.47	0.16	0.23	0.07	-0.06	0.27	0.26	-0.01	0.23	0.70
Marginal congestion costs												
Increased by 50%	0.37	0.36	-0.45	0.28	0.71	0.22	0	0.56	0.72	0.12	0.17	1.26
Reduced by 50%	0.17	0.31	-0.47	0.26	0	0.06	-0.31	0.47	0.05	-0.16	0.26	1.67
Value of wait time at transit stops												
Increased by 50%	0.30	0.48	-0.32	0.94	0.40	0.50	0.08	1.03	0.35	0.08	0.44	2.45
Reduced by 50%	0.24	0.19	-0.59	-0.33	0.31	-0.22	-0.32	0.07	0.37	-0.13	0.27	0.90
Vehicle size costs												
Increased by 50%	0.26	0.34	-0.65	0.29	0.31	0.15	-0.19	0.53	0.35	-0.02	0.32	1.60
Reduced by 50%	0.28	0.33	-0.25	0.26	0.40	0.13	-0.05	0.50	0.38	-0.02	0.39	1.58
Agency operating costs												
Increased by 50%	0.23	0.23	-0.51	-0.27	0.08	-0.36	-0.50	0.12	0.23	-0.24	0.24	1.35
Reduced by 50%	0.31	0.46	-0.39	1.18	0.70	0.79	0.42	1.13	0.56	0.40	0.47	1.84

Note

All values are in cents per passenger mile per one-cent increase in subsidy.

Table 5. Results for Marginal Welfare Effects with Alternative Assumptions for Agency Adjustment

	Washington				Los Angeles				London			
	Rail		Bus		Rail		Bus		Rail		Bus	
	Peak	Off-Peak	Peak	Off-Peak	Peak	Off-Peak	Peak	Off-Peak	Peak	Off-Peak	Peak	Off-Peak
$\varepsilon_V = 1.0$												
MW at current subsidy	0.20	0.27	-0.64	0.19	0.19	-0.14	-0.31	0.72	0.28	-0.38	0.11	2.41
marginal cost/price gap	-0.18	-0.42	-1.66	-2.39	-1.02	-1.90	-1.57	-1.68	-0.44	-0.96	-0.60	-0.52
net scale economy	0.12	0.56	0.83	2.35	0.25	1.21	0.75	1.89	0.05	0.31	0.40	2.65
externality	0.21	0.09	0.16	0.03	0.81	0.40	0.48	0.23	0.47	0.27	0.10	-0.51
other transit	0.06	0.04	0.03	0.20	0.14	0.15	0.03	0.28	0.19	-0.02	0.21	0.79
MW at 50% subsidy	0.20	0.29	0.11	0.45	0.37	0.43	0.22	0.82	0.31	-0.01	0.20	2.81
$\varepsilon_V = 0.5$												
MW at current subsidy	0.25	0.37	-0.36	0.18	0.44	0.28	-0.02	0.44	0.41	0.16	0.48	1.25
marginal cost/price gap	0.03	0.03	-0.64	-0.77	-0.39	-0.71	-0.60	-0.46	-0.10	-0.24	-0.10	0.13
net scale economy	0.00	0.28	0.14	0.94	-0.03	0.60	0.12	0.73	-0.05	0.15	0.04	1.00
externality	0.20	0.06	0.15	-0.07	0.78	0.30	0.45	0.04	0.46	0.25	0.47	-0.06
other transit	0.03	-0.01	0.00	0.08	0.08	0.08	0.01	0.13	0.09	-0.07	0.07	0.18
MW at 50% subsidy	0.25	0.37	-0.08	0.30	0.36	0.28	0.06	0.51	0.38	0.23	0.47	1.70

Note

All values are in cents per passenger mile per one-cent increase in subsidy.